



**HAL**  
open science

# Identification and validation of novel genes implicated in neurosensory and neurological diseases

Ariane Kröll-Hermi

► **To cite this version:**

Ariane Kröll-Hermi. Identification and validation of novel genes implicated in neurosensory and neurological diseases. Human health and pathology. Université de Strasbourg; Karlsruher Institut für Technologie, 2019. English. NNT : 2019STRAJ119 . tel-03851427

**HAL Id: tel-03851427**

**<https://theses.hal.science/tel-03851427>**

Submitted on 14 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse en cotutelle entre :

**ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE (ED414)**  
Laboratoire de Génétique Médicale U1112 Inserm, Université de Strasbourg

&

**KIT-FAKULTÄT FÜR CHEMIE UND BIOWISSENSCHAFTEN**  
Institute of Toxicology and Genetics, Karlsruher Institut für Technologie

# THÈSE

présentée par :

**Ariane KRÖLL-HERMI**

soutenue le : 19 decembre 2019

pour obtenir le grade de : **Docteur de l'université de Strasbourg/ Dr. rer. nat.**

Discipline/ Spécialité : Aspects Moléculaires et Cellulaires de la Biologie

## Identification and Validation of novel genes implicated in Neurosensory and Neurological Diseases

**THÈSE dirigée par :**

**Mme DOLLFUS Hélène**

**M WEISS Carsten**

Pr, Université de Strasbourg, Strasbourg

Dr, Karlsruher Institut für Technologie, Karlsruhe

**RAPPORTEURS :**

**Mme ZEITZ Christina**

**Mme LORENZ Birgit**

Dr, Institut de la Vision, Paris

Pr, Justus-Liebig Universität Giessen

---

**AUTRES MEMBRES DU JURY :**

**Mme KAPLAN Josseline**

**Mme GOLZIO Christelle**

**M SCHLUNCK Günther**

**M STRÄHLE Uwe**

Dr, Children's Hospital, Paris

Dr, Université de Strasbourg, Strasbourg Pr,

Albert-Ludwigs-Universität, Freiburg

Pr, Universität Heidelberg, Heidelberg



# Identification and Validation of novel genes implicated in Neurosensory and Neurological Diseases

Zur Erlangung des akademischen Grades einer

DOKTORIN DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

von der KIT-Fakultät für Chemie und Biowissenschaften

des Karlsruher Instituts für Technologie (KIT)

und

der École Doctorale 414 - Science de la Vie et de la Santé der Universität

Strasbourg (UdS)

genehmigte

DISSERTATION

von

**Ariane Kröll-Hermi**

1. Referentin: Prof. Hélène Dollfus
2. Referent: Dr. Carsten Weiss
3. Referent: Prof. Dr. Uwe Strähle
4. Referentin: Dr. Christina Zeitz
5. Referent: Prof. Dr. Birgit Lorenz
6. Referentin: Dr. Josseline Kaplan
7. Referentin: Dr. Christelle Golzio
8. Referent: Prof. Dr. Günther Schlunck

**Tag der mündlichen Prüfung : 19.12.2019**





# Preface

The following thesis was written in the context of a “Cotutelle de Thèse” co-directed by Hélène Dollfus head of the Laboratory of Medical Genetics UMRS\_1112 at the University of Strasbourg and Uwe Strähle/ Carsten Weiss of the Institute of Toxicology and Genetics at the Karlsruher Institute of Technology. The thesis was funded by the *Initiative d’Excellence* (Idex) program of the University of Strasbourg and the french-german university (UFA/DFH).

This thesis is based on the following research manuscripts:

**(1) Ariane Kröll-Hermi et al. (2019) Biallelic variations in *PRMT9* delineate a novel syndromic form of intellectual.**

***[Manuscript in preparation, will be submitted to the “American Journal of Human Genetics”.]***

**(2) Ariane Kröll-Hermi et al. (2019) “*MYH10* dominant variant causes Baraitser-Winter cerebro-fronto-facial syndrome related phenotype with major ophthalmic developmental features”**

***[Manuscript submitted to the “Journal of Medical Genetics”]***

**(3) Ariane Kröll-Hermi et al. (2019) “Proteasome subunit PSMC3 variants cause neurosensory syndrome combining deafness and cataract due to proteotoxic stress”.**

***[Manuscript submitted to the journal “EMBO Molecular Medicine”]***

# Table of Contents

1.	Introduction.....	1
	<b>1.1 Neurodevelopmental Disorders .....</b>	<b>2</b>
	1.1.1 Intellectual Disability .....	2
	1.1.2 Autism .....	5
	<b>1.2 Neurosensory Disorders .....</b>	<b>8</b>
	1.2.1 Eye diseases.....	8
	1.2.2 Early onset Cataract.....	9
	1.2.3 Coloboma of the eye .....	12
	1.2.4 Congenital Deafness .....	13
	<b>1.3 Identification of novel disease genes using Next-Generation Sequencing approaches .....</b>	<b>15</b>
	1.3.1 Whole Exome Sequencing .....	16
	1.3.2 Whole Genome Sequencing .....	16
	<b>1.4 Validation of Candidate Genes using the Zebrafish model .....</b>	<b>18</b>
	1.4.1 Zebrafish models for neurodevelopmental disorders.....	19
	1.4.2 Zebrafish models of neurosensory diseases .....	21
2.	Aim of the thesis.....	25
3.	Materials and Methods.....	27
	<b>3.1 Materials .....</b>	<b>27</b>
	<b>3.2 Methods carried out in the Laboratory of Medical Genetics.....</b>	<b>34</b>
	3.2.1 Patient Selection.....	34
	3.2.2 Genetic analysis .....	35
	3.2.3 Cell culture techniques .....	38
	3.2.4 Molecular biology techniques.....	38
	3.2.5 Biochemistry techniques .....	41
	<b>3.3 Methods carried out in the Institute of Toxicology and Genetics, Karlsruher Institut für Technologie (KIT).....</b>	<b>45</b>
	3.3.1 Zebrafish ( <i>Danio rerio</i> ) handling .....	45
	3.3.2 Generation of CRISPR/Cas9 knockout mutants .....	46
	3.3.3 Molecular biology methods .....	48
	3.3.4 Histological methods .....	51
	3.3.5 Behavior test of adult zebrafish.....	54
4.	Results .....	55
	<b>4.1 Project 1 .....</b>	<b>55</b>
	4.1.1 Synopsis.....	55
	4.1.2 Introduction – Project 1.....	58

4.1.3	Results – Project 1 .....	60
4.1.4	Discussion – Project 1 .....	74
4.1.5	Supplementary Data – Project 1 (S1) .....	78
<b>4.2</b>	<b>Project 2 .....</b>	<b>100</b>
4.2.1	Synopsis.....	100
4.2.2	Introduction – Project 2.....	102
4.2.3	Results – Project 2 .....	103
4.2.4	Discussion – Project 2 .....	110
4.2.5	Supplementary Data -Project 2 (S2).....	113
<b>4.3</b>	<b>Project 3 .....</b>	<b>121</b>
4.3.1	Synopsis.....	121
4.3.2	Introduction - Project 3.....	123
4.3.3	Results - Project 3 .....	124
4.3.4	Discussion - Project 3.....	136
4.3.5	Supplementary Information – Project 3 (S3) .....	140
<b>5</b>	<b>General Discussion.....</b>	<b>157</b>
<b>5.1</b>	<b>Advantages and limitations researchers are facing using WES or WGS for genetic diagnosis</b> .....	<b>157</b>
<b>5.2</b>	<b>Zebrafish as a model organism to study human heredity diseases – advantages and</b> <b>limitations.....</b>	<b>159</b>
<b>6</b>	<b>Conclusion and Perspectives.....</b>	<b>163</b>
<b>7</b>	<b>Résumé en français.....</b>	<b>164</b>
<b>8</b>	<b>References.....</b>	<b>175</b>
	<b>Curriculum Vitae .....</b>	<b>191</b>
	<b>Publications .....</b>	<b>192</b>
	<b>Talks/ Poster presentations .....</b>	<b>193</b>

Dedicated to my family



## Summary

Worldwide, about 5-10 % of the population suffer from rare genetic diseases. Although the interest in rare diseases has grown in recent years, there are often no therapies available. The reason for this, is especially that the genes responsible for rare diseases are still unknown. In order to get a better understanding of the molecular processes behind those diseases and to be able to find new therapeutic targets, it is very important to identify these yet unknown genes.

During my thesis I was particularly interested in the identification of disease genes associated with neurodevelopmental and neurosensory diseases. The strategy I used to identify new disease-causing genes is based primarily on the analysis of high-throughput sequencing data from patients, the confirmation of variants by Sanger sequencing and the performance of functional experiments using patients' cells and the zebrafish model to proof the pathogenicity of the novel gene mutations. Using this strategy, I was able to contribute to the identification of three novel disease genes.

In **project 1**, I report on 13 individuals with an Intellectual Disability (ID) syndrome caused by variations in the gene *PRMT9*. Beside ID, patients present mainly a global developmental delay, autism, epilepsy and hypotonia. By performing transcriptomic analysis, I was able to show that genes associated with ID and autism are differentially expressed in patients' fibroblasts. Moreover, a longer primary cilium length as well as differentially expressed cilia genes in patients' cells suggest a potential role of *PRMT9* during ciliogenesis.

In **project 2**, I contributed to the identification of a heterozygous missense mutation in the gene *MYH10* in three family members with autosomal dominant eyeball malformations, including coloboma, ptosis and craniofacial features. Given the symptoms a Baraitser-Winter cerebro-fronto-facial syndrome (BWCF) was suggested. Interestingly, *MYH10* is thought to interact with the only known BWCF proteins, *ACTB* and *ACTG1* and has been shown previously to be involved in the reorganization of the actin cytoskeleton. Indeed, in patients' cells, we observed an abnormal actin network. In zebrafish, a reduced expression of *myh10* resulted in eye and muscle anomalies.

In **project 3**, Whole-genome sequencing was performed on patients with severe deafness and early-onset cataracts as part of a neurological, sensorial and cutaneous novel syndrome. A unique homozygous variant in intron 10 of the *PSMC3* gene, encoding the 26S proteasome ATPase ring subunit 5, was identified with a predicted local splice effect as a new donor site. Functional experiments in patients' fibroblasts indicated that these cells are unable to compensate for proteotoxic stress. Two different *PSMC3* loss of function studies in zebrafish led to inner ear development anomalies as well as cataracts similar to the phenotype observed in patients.

This new finding can be an important step towards a better understanding of the pathophysiology of neurodevelopmental disorders and could possibly aid the search for therapeutic targets and pathways.

# Zusammenfassung

Weltweit sind etwa 5-10 % der Bevölkerung von einer seltenen genetischen Erkrankung betroffen. Obwohl das Interesse an seltenen Krankheiten in den letzten Jahren gewachsen ist, gibt es häufig noch keine Therapiemöglichkeiten. Um die molekularen Prozesse dieser Krankheiten besser zu verstehen und neue therapeutische Angriffspunkte zu finden, ist es daher sehr wichtig, alle Krankheitsverursachende Gene dieser Krankheiten zu identifizieren.

Im Verlauf meiner Doktorarbeit habe ich mich insbesondere auf die Identifizierung neuer Krankheitsgene konzentriert, die mit neurologischen- und neurosensorischen Erkrankungen assoziiert sind. Die Strategie, mit der ich diese Gene identifiziert habe, basierte hauptsächlich auf der Analyse von *Next-Generation Sequencing* Daten von Patienten, die Validierung der Varianten *via* Sanger-Sequenzierung und die Durchführung funktioneller Experimente unter Verwendung von Patientenzellen und dem Zebrafischmodell um die Pathogenität der neuen Genmutationen zu überprüfen. Mit Hilfe dieser Strategie trug ich zur Identifizierung drei neuer Krankheitsge bei.

In **Projekt 1**, wurden in 13 Individuen mit einer syndromalen Form der Geistigen Behinderung (mit Autismus, Epilepsie und Hypotonie) Varianten im Gen *PRMT9* identifiziert. Tatsächlich konnte ich feststellen dass viele Gene assoziiert mit Autismus und Geistiger Behinderung in den Zellen von Patienten in ihrer Transkription beeinträchtigt waren. Darüber hinaus deuten eine längere primäre Zilienlänge sowie die Beeinträchtigung der Transkription verschiedener Ziliengene in den Zellen dreier Patienten auf eine mögliche Rolle von *PRMT9* während der Entwicklung der primären Zilie hin.

In **Projekt 2** trug ich zur Identifizierung einer heterozygoten Missense-Mutation im Gen *MYH10* in drei Familienmitgliedern mit einem Baraitser-Winter-Cerebro-Fronto-Facial-Syndrom (BWCF) assoziiert mit Fehlbildungen des Augapfels (Kolobom, Ptosis und kraniofaziale Merkmale) bei. In vorangehenden Publikationen wurde gezeigt, dass *MYH10* *in silico* mit *ACTB* und *ACTG1* (BWCF assoziierte Gene) interagiert und an der Reorganisation des Aktin-Zytoskeletts beteiligt sein. Tatsächlich konnten wir in den Zellen der Patienten eine abnormale Struktur des Aktin-Cyotskeletts feststellen. Bei Zebrafischen führte eine verringerte Expression von *myh10* zu Augen- und Muskelanomalien.

In **Projekt 3** wurde eine Genomsequenzierung bei drei Familienmitgliedern mit schwerer Taubheit und früh einsetzendem Katarakt durchgeführt. Eine homozygote Variante in Intron 10 des *PSMC3*-Gens, die zu einem lokalen Spleißeffekt führt wurde identifiziert. *PSMC3* ist Teil des 26S-Proteasoms und wichtig für die Zellhomeostase. Tatsächlich wiesen Patientenzellen eine beeinträchtigte Proteinhomeostase auf und waren nicht in der Lage, proteotoxischen Stress zu kompensieren. Bei Zebrafischen führte eine reduzierte *psmc3* Expression, zu Anomalien der Innenohrentwicklung.

Die in dieser Arbeit gewonnen Ergebnisse tragen dazu bei ein besseres Verständnis für die Pathophysiologie dieser Krankheiten zu bekommen und helfen möglicherweise in Zukunft dabei neue Therapiemöglichkeiten zu entwickeln.



# Acknowledgements

At this point I would like to take the chance to thank all those who supported me during my doctoral thesis and have thus contributed to the success of this work.

First of all, my special thanks go to professor **Hélène Dollfus**, who fulfilled my dream to finally work in medical genetics by accepting my request to work in her laboratory as Master student and later as PhD student. My happiness and appreciation for this opportunity are unimaginable. Moreover, I would like to thank her for her scientific advices and feedback during the last three years.

In addition, I am deeply grateful to **Uwe Strähle** for accepting the “Cotutelle de Thèse” between his laboratory and my laboratory in Strasbourg and his useful advices during my time in Karlsruhe. I would also like to thank **Carsten Weiss**, without whom this Cotutelle would not have been possible. Thank you especially for your suggestions and feedback during the KIT seminars.

I also would like to thank the members of my thesis committee **Christina Zeitz**, **Birgit Lorenz**, **Christelle Golzio**, **Josseline Kaplan** and **Günther Schlunck** for accepting to read my manuscript and to evaluate my work even if I had to postpone the day of my defense. I am very grateful that, despite your busy schedules, you take your time to come to Strasbourg for the day of my defense.

My greatest thanks go to **Jean**, for his guidance, many scientific advices, his encouragement and his confidence in me. I cannot thank you enough for your supervision! During my entire PhD I was always able to count on your support!

A big thank you goes also to **Christelle** for her guidance, support throughout my whole project, for her inspirational advices, but also for the great conversations during lunch and for the sporty time in the fitness course of the KIT. I will miss those moments and wish you all the best!

In addition, I would like to thank **Corinne** for her encouragement, technical advices and for being such a nice person. Moreover, I especially would like to thank **Aline** for her help in the lab.

Furthermore, many thanks for the members of my lab and our neighbor group in Strasbourg with whom I also spend a lot of great moments apart from the lab **Daniel**, **Clarisse**, **Agnès**, **Katja**, **Jéremie** and **Géraldine**. Moreover, I would like to thank **Cathy**, **Véronique**, **Sophie** and **Elise** for their help and support.

Also, at the ITG, I had wonderful colleagues. I especially thank **Luisa**, **Sabrina**, **Swarnima**, **Vanessa**, **Gao**, **Sanamjeet**, **Amal**, **Angel**, **Amra**, **Irina**, **Helen** and **Lisa** for their encouragement, help, their good mood in the lab and for the great time we spend together after working hours. In addition, I would like to thank **Sepand**, **Masa**, **Volker** and **Victor** for many helpful scientific advices. I also wish to thank **Tanja**, **Anne** and **Silvia** for the funny conversations during lunch breaks.

Finally, I would like to thank my **parents** and my sister **Delia** who supported me in many ways. And most of all, I thank my husband **Mehdi** for his love, support and encouragement. There are no words to express my gratitude!

# List of figures

Figure 1. IQ distribution in the population and ID severity levels.....	2
Figure 2. Illustration of the human eye and retina anatomy. ....	8
Figure 3. Development of the lens in humans. ....	10
Figure 4. In patients affected by coloboma, the optic fissure did not close properly during the embryonal eye development. ....	12
Figure 5. Schematic representation of the human ear and cross section of the cochlear.....	14
Figure 6. Four main steps of Next Generation Sequencing. ....	15
Figure 7. Gene knockdown versus knockout in zebrafish.....	19
Figure 8. Comparison between the human and zebrafish brain.....	20
Figure 9. Comparison of the global structure of the human and zebrafish eye. ....	21
Figure 10. Anatomical comparison between the human inner ear and the embryonic zebrafish ear. ....	23
Figure 11. Strategy used to identify novel disease genes and my personal contribution to the three projects of my thesis. ....	25
Figure 12. Identification of a loss of function mutation in the gene PRMT9 in the index family A. ....	62
Figure 13. Effect of the PRMT9 variation p.Leu182Trpfs* on the mRNA and protein expression. ....	63
Figure 14. PRMT9 mutations found in nine families. ....	64
Figure 15. Mutations in the genes PRMT9 affect the cilia length in patient fibroblasts. ....	66
Figure 16. In affected individuals, PRMT9 fails to methylate the splicing factor SAP145. ....	67
Figure 17. Multiple genes associated with ID, autism or cilia biogenesis/ function are dysregulated in individuals carrying biallelic PRMT9 mutations.....	69
Figure 18. MZprmt9 <sup>-/-</sup> mutants do not display a typical ciliopathy phenotype, but show behavioral abnormalities. ....	71
Figure 19. Personal contribution to project 2. ....	101
Figure 20. Pictures of the patients.....	104
Figure 21. Three affected family members carry a heterozygous MYH10 variant. ....	105
Figure 22. MYH10 is less expressed and mis-localized in affected fibroblasts than control (ctrl) cells, affecting actin filaments length and cytoskeleton dynamics.....	106
Figure 23. Knock-down of myh10 affects the eye development and the organization of skeletal muscle fibers in 3-day-old zebrafish embryos.....	108
Figure 24. Personal contribution to project 3. ....	122
Figure 25. Family pedigree with three affected individuals.....	124
Figure 26. Mutation segregation analysis of PSMC3 intron 10 mutation.....	126
Figure 27. cDNA analysis. ....	127
Figure 28. Effect of the deep intronic PSCM3 variant on the proteasome function.....	128
Figure 29. Fibroblasts derived from patient carrying the c.1127+337A>G homozygous PSMC3 variation exhibit an increased amount of both proteasome complexes and ubiquitin-protein conjugates. ....	130
Figure 30. Patient fibroblasts carrying the c.1127+337A>G homozygous PSMC3 variation exhibit an exhausted TCF11/Nrf1 processing pathway which prevents them to upregulate proteasome subunits in response to proteotoxic stress.....	131
Figure 31. psmc3 morphants and F0 mosaic zebrafish exhibit cataract and show abnormalities during the semicircular canal development in the ear. ....	133

## List of supplemental figures

Figure S1.1 Variation filtration from whole exome sequencing data of the index patient (A.II-1) and identification of possible pathogenetic variants using VaRank.....	78
Figure S1.2 Exclusion of potential candidate genes via Sanger Sequencing. ....	78
Figure S1.3 Multiple protein sequences alignment of human PRMTs together with metazoan PRMT9 .....	79
Figure S1.4 Biallelic <i>PRMT9</i> variations affect the alternative splicing of genes.....	79
Figure S1.5 Knockout of <i>prmt9</i> in zebrafish using the CRISPR/Cas9 system. ....	80
Figure S1.6 No differences during brain development were observed in <i>MZprmt9</i> <sup>-/-</sup> zebrafish embryos.....	81
Figure S1.7 Morpholino efficiency of <i>prmt9</i> -mo was assessed by co-injecting a <i>prmt9</i> N-terminal sequence fused to GFP. ....	81
Figure S1.8 <i>prmt9</i> -mo injected embryos present a curved body axis (3dpf) and an impaired brain morphogenesis (24 hpf).....	82
Figure S2.1 Actin filaments are shorter in patient's skin fibroblasts.....	114
Figure S2.2 Optic fissure closure is delayed in <i>myh10</i> morphants.....	114
Figure S2.3 Morpholino efficiency test of <i>myh10</i> -morphants. ....	115
Figure S2.4 A second morpholino ( <i>myh10</i> - <i>mo2</i> ) confirmed the eye and muscle phenotype. ....	115
Figure S2.5 Generation of a <i>myh10</i> KO mutant using CRISPR/Cas9.....	116
Figure S2.6 <i>myh10</i> <sup>-/-</sup> mutants do not present any obvious phenotype. ....	117
Figure S3.1 Analysis of homozygosity region in chromosome 11.....	142
Figure S3.2 Selection of candidate genes related to deafness and/or cataract. ....	143
Figure S3.3 Types of interactions between CHMP4B, ACTG1, GJB6 and PSMC3.....	143
Figure S3.4 Identification of the boundaries of the inserted cryptic exon.....	144
Figure S3.5 <i>PSMC3</i> mRNA quantification.....	144
Figure S3.6 No difference in <i>PSMC3</i> expression or localization.....	145
Figure S3.7 <i>In situ</i> hybridization with <i>psmc3</i> antisense probe showing ubiquitous expression (4 hpf-48 hpf) .	146
Figure S3.8 <i>psmc3</i> morphants and crispants exhibit smaller lenses and inner ears.....	146
Figure S3.9 Otolith development is not affected in <i>psmc3</i> morphants.....	147
Figure S3.10 <i>psmc3</i> morphants display no obvious brain malformations at 24 hpf.....	148
Figure S3.11 A second guide RNA ( <i>sgRNA1</i> ) confirms the cataract and ear phenotype seen in morphants ( <i>mo</i> ) and crispants ( <i>sgRNA2</i> ).....	149
Figure S3.12 Expression of genes involved in the zebrafish inner ear development in <i>psmc3</i> crispants. ....	150
Figure S3.13 Design and efficiency of morpholinos targeting <i>psmc3</i> pre-mRNA and guide RNAs targeting <i>psmc3</i> .....	151

## List of tables

Table 1. Classification of different severity levels of intellectual developmental disorders according to ICD-11. .	3
Table 2. Comparison of the advantages and disadvantages of WES and WGS. ....	17
Table 3. Oligonucleotides.....	27
Table 4. Oligos ordered for the synthesis of CRISPR guide RNAs and the corresponding primers used for genotyping. ....	29
Table 5. Plasmids used in this study. ....	29
Table 6. In situ hybridization probes used in this study. ....	30
Table 7. Zebrafish lines used in this study. ....	30
Table 8. Morpholinos used in this study. ....	31
Table 9. Antibodies used in this study. ....	31
Table 10. Solutions and buffers used in this study. ....	32
Table 11. Software used in this study. ....	33
Table 12. My personal contribution to project 1. ....	57
Table 13. Summary of the Major Clinical Features of Patients with biallelic Pathogenic Variants in PRMT9 .....	61
Table 14. Distribution of Clinical characteristics with the PRMT9 cohort. ....	65
Table 15 Enrichment in Genes Involved or Potentially Involved in ID, ASD or cilia function/biogenesis. ....	68
Table 16. Main dysmorphic features of the 3 affected individuals and observed in actinopathies. ....	103
Table 17. Clinical description of the patients with PSMC3 pathogenic variants.....	125

## List of supplemental tables

Table S1.1 Homozygous regions specific to the affected individual A.II-1. ....	83
Table S1.2 Differentially expressed genes (+FCS).....	83
Table S1.3 Functional Annotation (+FCS).....	85
Table S1.4 Differentially expressed genes (-FCS/ ciliary conditions).....	87
Table S1.5 Functional Annotation (-FCS/ ciliary conditions).....	93
Table S1.6 Alternative Splicing (+FCS).....	96
Table S1.7 Alternative Splicing (-FCS/ciliary conditions).....	97
Table S2.1 List of 2 239 connected human genes to ACTB and/or ACTG1 provided by the FunCoup 4.0 human database.....	118
Table S2.1 Summary of the whole exome sequencing results. No potentially pathogenic variant was identified in ACTB and ACTG1 genes.....	120
Table S3.1 Summary of the whole exome sequencing results.....	152
Table S3.2 Summary of the whole genome sequencing results. ....	153
Table S3.3 Orthologous ID equivalent of PSMC3, ACTG1, CHMP4B and GJB6 in human, mouse and yeast. ....	154
Table S3.4 Publications showing genes interactions between PSMC3, ACTG1, CHMP4B and GJB6. ....	154
Table S3.5 Proteasome proteins identified using nanoLC-MS/MS analysis and quantified by Spectral Count. .	155

## List of Abbreviations

ADMA	Asymmetric dimethylarginine	ID	Intellectual Disability
ADHD	attention deficit hyperactivity disorder	IQ	intelligence quotient
ASD	autism-spectrum disorder	ICD	International Statistical Classification of Diseases and Related Health Problems
AGC	automatic gain control	LoF	loss of function
BWCFF	Baraitser Winter Cerebro Fronto Facial syndrome	MACS	magnetic micro particles
BBS	Bardet Biedl Syndrome	MS	mass-spectrometry
CAM	Cell-adhesion molecule	MZ	maternal zygotic
CNS	central nervous system	MMA	Monomethylarginine
CARGO	Centre de référence pour les Affections Rares en Génétique Ophtalmologique	NGS	Next Generation Sequencing
CNRGH	Centre National de Recherche sur le Génome Humain	NMD	Nonsense-mediated decay
CGH	comparative genomic hybridization	PSM	peptide spectrum matches
DDA	data-dependent acquisition mode	PVNH	Periventricular nodular heterotopias
dpf	day post fertilization	PKU	phenylketonuria
DD	Developmental Delay	PCR	Polymerase chain reaction
ENT	Ear, Nose and Throat	PRMT	Protein arginine methyltransferase
FCS	Fetal calf serum	PAM	Protospacer Adjacent Motif
FISH	Fluorescence <i>in situ</i> hybridization	qPCR	Quantitative polymerase chain reaction
FC	Fold change	RT	Reverse transcript
GO	gene ontology	sgRNA	Single guide RNA
GFP	Green fluorescent protein	SNP	Single nucleotide polymorphism
gRNA	Guide RNA	SNV	Single nucleotide variant
hpf	hours post fertilization	SV	Structural Variations
HYB	hybridization buffer	SDMA	Symmetric dimethylarginine
IF	Immunofluorescence staining	TRP	Tetratricopeptide repeats
IP	immunoprecipitation	UPS	Ubiquitin Proteasome System
ISH	<i>in situ</i> hybridization	VUR	Vesicoureteral Reflux
IDD	intellectual developmental disorder	WES	Whole Exome Sequencing
		WGS	Whole Genome Sequencing
		WT	Wilde type
		WHO	World Health Organization



# Introduction

# 1. Introduction

Thanks to next-generation sequencing (NGS) approaches, countless new disease genes have been identified over the last decade. The identification of these genes is essential to understand the physiological role of a gene/protein, to get a better knowledge about the pathophysiology of a disease, to find new targets for therapies, and to provide genetic counseling to families of affected relatives. Although the way from the identification of a new disease gene to the development of appropriate therapies is very long, so that the families might not profit from a therapy before decades, with the identification of the genetic trigger, families can better assess the level of risk for another pregnancy or for another family member. In addition, the course of the disease can be better estimated and the families can participate in special patient groups giving them the feeling of being supported and not being alone.

The aim of this “Cotutelle de Thèse” co-directed by H el ene Dollfus head of the Laboratory of Medical Genetics UMRS\_1112 at the University of Strasbourg and Uwe Str ahle head of the Institute of Toxicology and Genetics at the Karlsruher Institute of Technology, was the identification of novel genes associated with neurodevelopmental or neurosensory rare diseases.

In the past few years, many successful collaborations between the two laboratories have led to the identification of novel disease genes, in particular for a group of diseases called ciliopathies.<sup>1-4</sup> The laboratory of medical genetics UMRS\_1112 works in close collaboration with the service of medical genetics and the reference center for genetic ophthalmologic diseases (CARGO) at the Strasbourg University hospital allowing them a direct contact with patients and ad hoc recruitment for research . In addition, bioinformatics analysis and experiments on patients’ cells have been carried out in the laboratory UMRS\_1112 Strasbourg, whereas appropriate zebrafish models for the respective candidate genes have been generated at the ITG in Karlsruhe.

Thus, during my PhD thesis, I had the opportunity to benefit from the experience and the help of both laboratories, which allowed me to contribute to the identification of three novel disease genes associated with neurodevelopmental or neurosensory diseases.

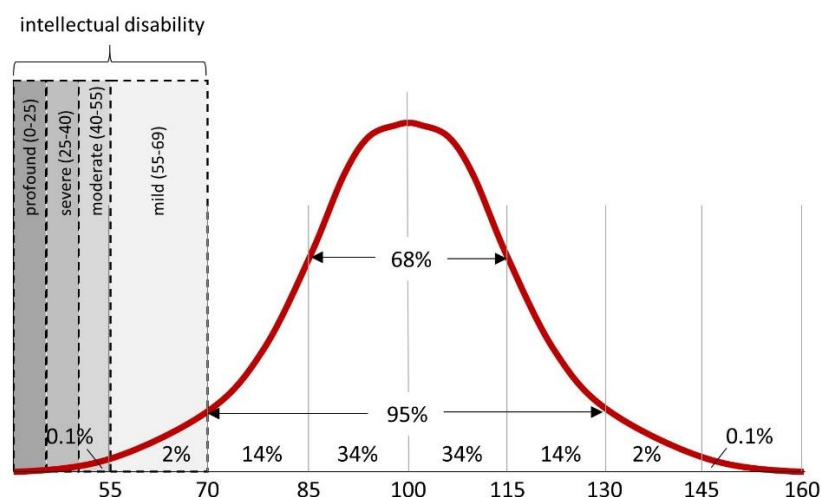
## 1.1 Neurodevelopmental Disorders

Neurodevelopmental disorders are a group of diseases associated with an abnormal neuronal development and behavior. The term neurodevelopmental disorder includes for example, Intellectual disability (ID), developmental delay (DD), autism-spectrum disorder (ASD) or language disorders. Worldwide, about 3 % of all children are affected by a neurodevelopmental disorder.<sup>5,6</sup> With more than 1000 associated loci, neurodevelopmental diseases are extremely heterogeneous.<sup>6</sup>

### 1.1.1 Intellectual Disability

#### General description of Intellectual Disability

Intellectual Disability (ID), previously called mental retardation, is associated with an impaired cognitive and adaptive development, affects 1-3 % of individuals worldwide and is one of the most important healthcare problems.<sup>7</sup> In the Netherlands, ID was even ranked as one of the top 20 most costly disorders.<sup>8</sup> The cognitive impairment of patients with intellectual disabilities occurs before the age of 18, affects the development and has lifelong consequences.<sup>7</sup> The severity of ID is determined by the intelligence quotient (IQ). Based on the IQ, the ICD-10 (International Classification of Disease 10<sup>th</sup> Revision) published by the World Health Organization (WHO), classifies the severity into four categories: An IQ between 50-55 to approximately 70 is defined as mild ID. Patients with an IQ between 35-40 to 50-55 have a moderate form of ID. A severe form is diagnosed when the IQ is between 20-25 to 35-40. An IQ less than 20-25, is considered a profound form of ID (Figure 1).



**Figure 1. IQ distribution in the population and ID severity levels.**

(inspired by Johns, 2014)<sup>9</sup>

This year (May 2019), however, a new version of the international classification was adopted, the ICD-11, and comes into effect on the first of January 2022. In this version, the term “intellectual disability” is replaced with “intellectual developmental disorders” (IDD). From 2022 on, the classification of IDD



severity levels will not only be based on standardized IQ tests, but also rely on a description of clinical characteristics (Table 1).

**Table 1. Classification of different severity levels of intellectual developmental disorders according to ICD-11.**

	<b>Language Skills</b>	<b>Practical Activities</b>	<b>Need of Support</b>
<b>6A00.0 mild IDD</b>	Difficulties in understanding and learning complex language skills/ academic skills	Most can take care of themselves and their domestic	Once adult, they can live and work independently with an adequate support
<b>6A00.0 moderate IDD</b>	Generally limited to basic skills	Some may take care of themselves and their domestic	Once adult, they can live and work independently but most will need a consistent support
<b>6A00.0 severe IDD</b>	Very limited language skills/ capacity to obtain academic skills	May have impaired motor skills, with a lot of exercise and support some may take care of basic self-care skills	require daily support/ supervision
<b>6A00.0 profound IDD</b>	Very limited language skills/ capacity to obtain basic academic skills	May have impaired motor and sensory skills, need support for adequate care	require daily support in a supervised environment

Table created using information of the World Health Organization - International Classification of Disease 11<sup>th</sup> Revision (mai 2019)

ID can be caused by non-genetic and genetic factors. Common non-genetic causes are for example prenatal (e.g. rubella, toxoplasmosis or herpes zoster) and postnatal infections (e.g. meningitis, Japanese encephalitis, and tuberculosis), premature birth and a traumatic impact (brain hemorrhages, lack of oxygen and serious head injuries).<sup>7</sup> Shockingly, 5 % to 20 % of ID is caused by nutritional deficiencies as well as alcohol and drugs abuse.<sup>7</sup> ID caused by genetic factors was reported in 30-50 % of the cases.<sup>10</sup> Especially, in the last decade many novel ID genes could be identified thanks to the rapid advance in NGS technologies. Seeing this development, Milani and her colleagues assumed that the proportion of identified genetic causes for ID, will soon increase from 50 % to 80 %.<sup>11</sup> Genetic forms of ID can be divided in two categories. Syndromic forms of ID, always occur accompanied by other symptoms, while patients with isolated ID do not present further clinical anomalies. Especially, the prevalence of additional neurodevelopmental features is high in individuals presenting ID. Epilepsy for example occurs in about 22.2 % of ID patients with a higher prevalence in severe cases.<sup>12</sup> Moreover, behavior anomalies such as non-compliance, property destruction, tantrums, meltdowns, and physical aggression towards others or self-directed appear in about 25 % of individuals diagnosed with ID.<sup>13</sup> Also, psychiatric disorders (ADHD, mood disorders, anxiety disorders, and psychotic disorders) have a

higher prevalence in ID patients than in the whole population.<sup>8</sup> In addition, 70 % of autism patients have been reported to be also affected by ID.<sup>14</sup> These results indicate that, Autism and ID genes are implicated in the same or overlapping signaling pathways.

### **Genetic causes of Intellectual Disability**

Chromosomal abnormalities are one of the main causes of ID. Especially, aneuploidy, which describes an abnormal number of chromosomes in cells, can be easily diagnosed with conventional cytogenetic approaches, such as the conventional karyotyping, high-resolution karyotyping, fluorescence in situ hybridization (FISH) or comparative genomic hybridization (CGH).<sup>15</sup> The most common chromosomal aberration found in ID patients is trisomy 21, which is the cause of Down syndrome (MIM #190685).<sup>16</sup> Other aneuploidies associated with ID are, trisomy 13 (Patau syndrome) or trisomy 18 (Edwards syndrome). These syndromes, however, are lethal in the first weeks of life.<sup>17</sup> In addition, also large (> 5Mb) deletions or duplications of chromosome segments can cause ID, e.g. the microdeletion 5p (Cri-du-Chat syndrome, MIM #123450) and the microdeletion 4p (Wolf-Hirschhorn syndrome, MIM #194190).<sup>18</sup> These can be identified via FISH or high-resolution karyotyping. Another 13 % of ID patients have submicroscopic aberrations (< 5 Mb), such as the 7q11.23 microdeletion (Williams-Beuren syndrome, MIM # 194050).<sup>16,19</sup> They can be detected performing a microarray comparative genomic hybridization (aCGH).

In about 9 % of ID patients, changes in individual genes have been identified as disease-causing.<sup>16</sup> For a long time the identification of monogenic disease genes was limited to familial forms with several affected individuals. Since NGS technologies are used for genetic diagnostics, however, a rapid increase of identified ID genes was observed.<sup>20</sup> This development allowed also the identification of many *de novo* mutations. To date, mutations in more than 700 genes have been reported in ID patients.<sup>21</sup> And despite the yearly increasing number of identified ID-genes, the curve has not yet reached its plateau, assuming that many ID genes have not yet been identified.<sup>20</sup>

Among the identified genes, over 100 are associated with X-linked ID.<sup>22</sup> It was estimated that X-linked ID accounts for approximately 10-20 % of inherited cases of ID.<sup>18</sup> The most common X-linked form of ID is the Fragile X syndrome (MIM # 300624), which is caused by trinucleotide repeats in the first intron of the *FMR1* gene.<sup>23</sup> Moreover, genetic mutations associated with ID can be transmitted according to an autosomal dominant or recessive mode.

Among the most commonly mutated genes, inherited through an autosomal dominant mode, are e.g. *ARID1B* (Coffin-Siris Syndrome, MIM #135900), *SCN2A* (Early infantile epileptic encephalopathy MIM #613721 or benign familial infantile seizures MIM #607745), *ANKRD11* (KBG syndrome, MIM #148050) or *SATB2* (Glass syndrome MIM #612313).<sup>24</sup> One of the most common forms of autosomal recessive

ID is phenylketonuria (PKU, MIM # 261600), which is also one of the few treatable forms of ID that can be improved by a diet low in phenylalanine.<sup>25</sup>

Overall, it was estimated that most of the dominant inherited genes are already identified, whereas an implication of more than 3000 recessive inherited genes is expected.<sup>26</sup>

### **Molecular Pathways involved in ID**

Up to date, over 700 ID genes have been identified and included in the SysID database.<sup>21</sup> Performing an enrichment analysis based on gene ontology (GO) terms, Kochinke and colleagues revealed an implication of those genes in metabolism, transport, nervous system development, RNA metabolism and transcription. Most enriched GO terms were the hedgehog pathway, glutamate receptor signaling, peroxisome, glycosylation and cilia.<sup>21</sup>

A more detailed description of pathway involved in ID in association with autism is given in the following chapter (**Pathways linked to autism in combination with ID**).

## 1.1.2 Autism

### **Definition of Autism**

Autism is a developmental disability belonging to a complex set of syndromes called autism spectrum disorders (ASDs). Beside autism, ASD includes e.g. Asperger syndrome or Rett syndrome, and is characterized by difficulties in social behavior and non-verbal communication, as well as repetitive and stereotype behaviors.<sup>27</sup> The guidelines for the diagnosis of ASD and the classification into the various syndromes are described in the Manual for the Diagnosis and Statistics of Mental Disorders (DSM-5). The exact diagnosis, however, is not always clear, as the phenotypes may also overlap. Worldwide, more than 1 % of children suffer from ASD, with a 4-5 times higher prevalence in males than in females.<sup>28</sup>

### **Causes of ASD**

Causes of ASD are genetic and environmental factors.<sup>29</sup> Multiple studies in twins and families, revealed a high heritability, suggesting that genetic factors play a major role in ASD.<sup>30-32</sup> Especially, common genetic variations, acting additively, contribute to the development of ASD.<sup>33,34</sup> Rare, novel or *de novo* variations, on the other hand, are major risk factors for the individual liability.<sup>33,35</sup> These studies also demonstrate, that the recurrence risk in families with one child affected with ASD is far below the expected 25 %, indicating that the inheritance of ASD does not comply with the inheritance of Mendelian monogenetic rules.<sup>31</sup> Nevertheless, the prevalence for a sibling to develop ASD is with 2-6 % higher than in the overall population (0.04 %).<sup>31</sup> ASD rather arises through the combination of different affected loci in the genome and environmental factors.<sup>29</sup> Although an association with autism

has been demonstrated for more and more genes and chromosomal regions, many genetic factors have not been identified yet.

### **Pathways linked to autism in combination with ID**

70 % of autism patients are also affected by ID, therefore it was suggested that they are likely to be biochemically and molecularly related. Most of the proteins associated to both ID and autism are expressed in the synapses and implicated in various cellular processes, that are mainly associated with the regulation of the neuronal morphology and communication.<sup>14</sup>

#### 1. Regulation of cytoskeleton dynamics

The cytoskeleton gives neurons its shape and stabilizes the entire cell. Moreover, it is implicated in axon outgrowth and synapse formation and thus important for remodeling of synaptic connections.<sup>36</sup> Interestingly, several proteins affected in ID and autism, have been shown to be implicated in pathways that regulate the organization of the actin-cytoskeleton and hence have an impact on the structure and function of dendrites and synapses.<sup>14,37</sup> These proteins are mainly proteins of the GTPase signaling pathway, such as RhoA, Rac and Cdc42. In addition, also regulators and effectors of these proteins have been associated with ID and autism: OPHN1, MEGAP, OCRL1, ARHGEF6, ARHGEF9, FGD1, LIMK1, PAK3, and IQSEC2.<sup>14,37</sup>

#### 2. Presynaptic vesicle cycling and exocytosis

The formation and function of chemical synapses is essential for the neuronal communication. Especially the transport, storage and release of neurotransmitters through synaptic vesicles play an important role for the synaptic communication. Proteins associated with the regulation of vesicle cycling and exocytosis at the synapse have been reported to be mutated in patients with ID and autism. Synapsins, Syn1 and Syn11, for example have been reported to help to maintain a reserve pool of vesicles and to be implicated in the post docking step of exocytosis. Mutations in *SYN1* and *SYN11* lead to a strong epilepsy and autism phenotype in human and/or mice.<sup>14,38,39</sup>

Another protein, RAB39B, encodes a protein involved in intracellular vesicle trafficking. Patients with mutations in *RAB39B* have a syndromic form of ID in association with epilepsy and autism.<sup>40</sup> Also mutations in genes encoding proteins involved in the fusion of vesicles with the presynaptic membrane (e.g. the SNARE protein Munc18-1) and implicated in exocytosis and transmitter release, such as *IL1RAPL1*, have been shown to be associated with autism, ID and/or epilepsy.<sup>41,42</sup>

### 3. Translational regulation, protein degradation and turnover

In addition, two quality control systems seem to play a role in synaptic plasticity, the ubiquitin proteasome system (UPS) and nonsense-mediated decay (NMD). The UPS is involved in the regulation of intracellular protein levels and protein quality control and thus plays an important role in cell homeostasis.<sup>43</sup> NMD, on the other hand, is a protective mechanism of the cell and is intended to prevent the synthesis of faulty proteins and gain of function effects, as the mechanism recognizes and degrades mRNA transcripts containing premature stop codons.<sup>44</sup>

Mutations and copy number variants in genes involved in NMD, such as, *UPF3B*, have been shown to cause ID and autism.<sup>45</sup> And also, mutations in UPS-associated genes, such as *UBE3A*, *HUWE1* and *UBR1* can lead to a combined ID and autism phenotype.<sup>14,46</sup>

### 4. Transcriptional regulation

Interestingly, also many genes implicated in the transcriptional regulation, such as transcription factors and chromatin modifiers, have been associated with ID, highlighting the importance of a precise temporal and spatial regulation of gene expression for the brain development.<sup>47</sup> For example, mutations in the *MECP2* gene, a gene encoding for a transcriptional repressor, cause Rett syndrome, a neurodevelopmental disorder associated with ID and autism.<sup>48</sup> Moreover, chromatin modifiers, such as *BRPF1*, a histone acetylase modifier was recently associated with ID.<sup>49</sup>

### 5. Cell adhesion molecules in trans-synaptic signaling (synaptic cleft)

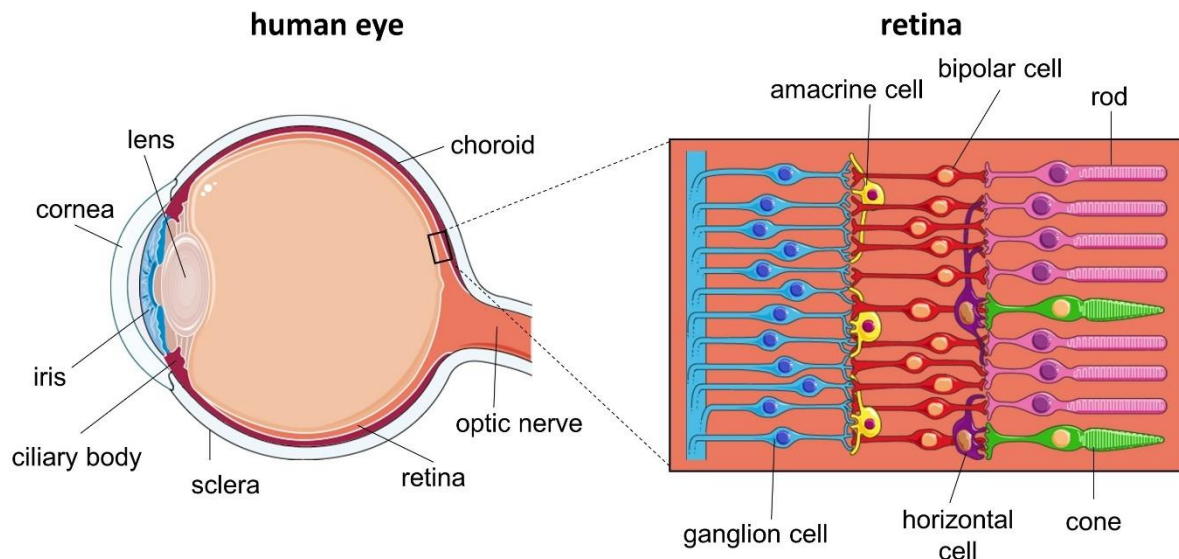
Cell-adhesion molecules (CAMs) play an important role initializing contacts between pre- and postsynapse, functional maturation and maintenance. The identification of rare genetic variations in CAMs, such as *PCDH10*<sup>50</sup>, *SHANK2*<sup>51</sup> and *SHANK3*<sup>52</sup>, described in humans and/ or mice with an ID/autism-like phenotype, strengthens this hypothesis.<sup>14</sup>

## 1.2 Neurosensory Disorders

Depending on the severity of symptoms and the time of onset of the disorder, a functional impairment of the sensory system, such as vision or hearing, can lead to serious social and cognitive problems. According to the WHO, worldwide 285 million people suffer from an impaired vision and 466 million from hearing impairment. Moreover, neurosensory diseases, such as cataract, represent a great burden for the health care system, as life expectancy increases and the number of affected individuals steadily rises.

### 1.2.1 Eye diseases

The human eye is a complex sensory organ that has to fulfill different requirements. It has to be movable, must be able to react quickly to new stimuli, enables sharp vision of close and distant objects, tracks movements and enables vision at day and night. The main function of the eye is to collect light, transform it into signals and transport it to the brain. In Figure 2 the main structures of the human eye are illustrated.



**Figure 2. Illustration of the human eye and retina anatomy.**

(This figure was drawn using the image bank of Servier Medical Art. Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License; inspired by Keeling et al., 2018)<sup>53</sup>

Basically, the human eye consists of three primary layers.<sup>54</sup> The sclera and cornea are the outer coat of the eye, have a protective function and are responsible to maintain the eye's shape.<sup>55</sup> Moreover, the transparent cornea is, as the lens, responsible to bend the rays of the incoming light and to focus the light to the retina.<sup>55</sup> The middle layer comprises the iris, ciliary body and choroid.<sup>54</sup> The amount of light entering the eye is controlled by the iris, which can adjust the size of the pupil depending on the

light condition. The ciliary body produces the aqueous humor, which is important to remove waste as well as to provide nutrients and oxygen to the lens and the cornea. In addition, the ciliary muscle is able to adjust the shape of the lens and thus improves the focus depending on the distance of an observed object.<sup>56</sup> The choroid provides the eye with oxygen and nutrients.<sup>57</sup> The inner layer, also called retina is a light-sensitive layer, composed of nerve- and photoreceptor cells. Humans have two types of photoreceptor cells: rods and cones. Rods are required for black and white vision and work better in dim light whereas cones are responsible for sharpness and the perception of colors in bright light. The light and color information received by the photoreceptors is subsequently transmitted *via* the associated nerve cells (bipolar and ganglion cells) and the optic nerve to the visual cortex. Horizontal and amacrine cells are responsible for the lateral interconnection within the retina and improve e.g. the contrast by comparing the information obtained by different photoreceptors.<sup>58</sup>

If any of those ocular structures does not work properly, vision is impaired. Common eye diseases are e.g. glaucoma, cataract, macular degeneration, myopia. More than 900 rare eye diseases exist (retinitis pigmentosa is one of the most common) and taken together represent the most common cause of severe visual impairment in Europe.<sup>59</sup> To get a better understanding of rare ocular diseases and to find new targets for therapies the identification of genes and their variants is essential. In the past years many eye disease genes have been identified. My laboratory in Strasbourg and Karlsruhe have contributed to the identification of many novel disease genes in particular associated with the Bardet-Biedl syndrome (BBS), a disorder associated with retinitis pigmentosa (RP), polydactyly, obesity, hypogonadism, kidney dysfunction and cognitive impairment.<sup>2,60-64</sup> RP is characterized by abnormalities of the photoreceptors or the retinal pigment epithelium and leads to a progressive vision loss.<sup>65</sup> The identification of BBS genes, genes implicated in the biogenesis and function of the primary cilium and the connecting cilium of photoreceptors, has led to an improvement of the understanding of the pathophysiology of RP.<sup>66</sup>

The objective of this thesis, however, is mainly to improve the understanding of ultra-rare eye diseases in syndromes associating congenital cataract/deafness and associating coloboma/ptosis by identifying the disease-causing genes in two affected families.

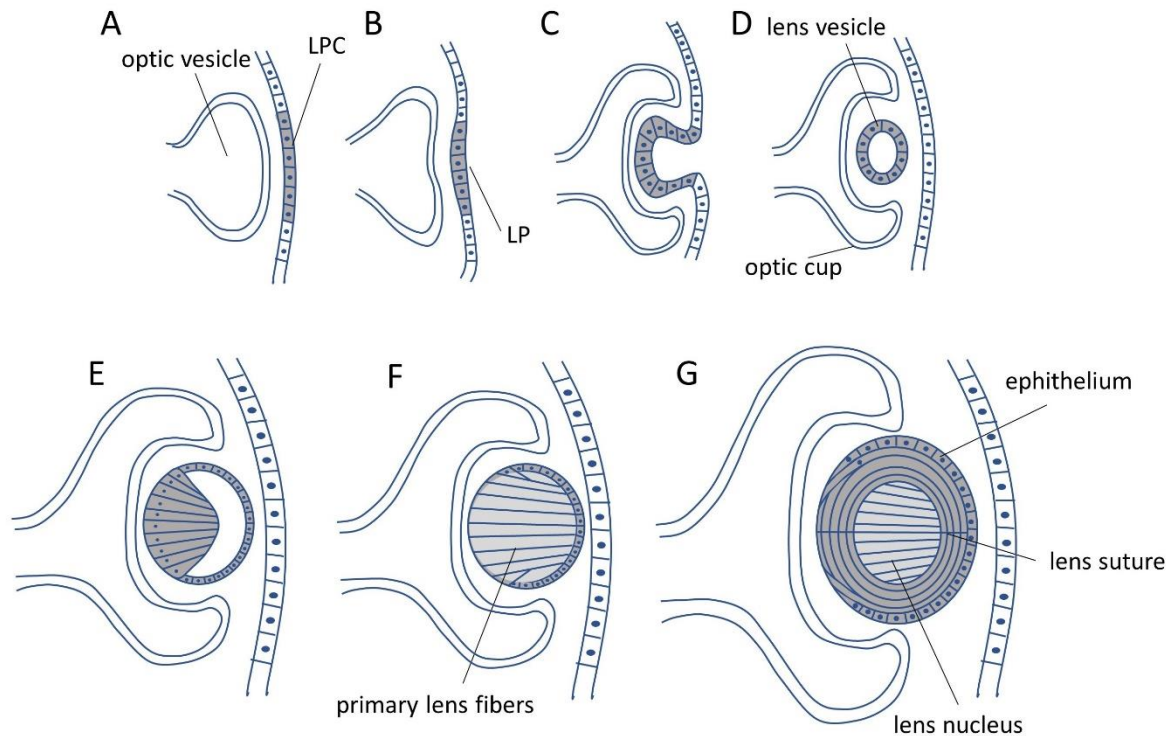
### 1.2.2 Early onset Cataract

#### **General description**

Congenital cataract is the opacity of the eye lens that is present from birth on and by extension to early-life cataracts. It is a rare disease with a varying prevalence between developed (1-6/10,000) and developing countries (5-15/10,000).<sup>67</sup> Nevertheless, congenital cataract is the most common cause of treatable childhood blindness.<sup>68</sup> A gender difference was not observed.<sup>69</sup> Cataracts can occur isolated

(non-syndromic), in conjunction with other eye abnormalities, or can be part of a multisystem disorder (e.g. Down syndrome).<sup>68</sup>

### Development of the lens



**Figure 3. Development of the lens in humans.**

(A) The lens develops from the ectoderm. (B-E) The ectoderm invades and forms the lens vesicle. (F-G) Cells on the posterior side of the lens form the primary lens fibers and thus the later lens nucleus. (H) Representation of the mature lens. Secondary fibers form around the core. The eye continues to grow for a lifetime. LPC: Lens Placode; LP: Lens Pit. (inspired by Cvekl and Zhang, 2017<sup>70</sup> and Richardson et al., 2017<sup>71</sup>)

The lens exclusively develops from the ectoderm, explaining also its involvement in generalized skin diseases. Eye development starts with the formation of the optic vesicle, followed by the thickening of the ectoderm leading to the formation of the lens placode and retinal disc. Invagination of the lens placode and optic vesicle initiates the formation of the lens vesicle and the optic cup. The lens vesicle detaches from the optic cup and the cells forming the lens fibers start to elongate, until they fill the whole lens vesicle. Once they contact the anterior lens vesicle, an epithelium is formed. This structure represents the future core of the adult lens. The inner layer of the optic cup begins to form the retina. The secondary lens fibers, which are arising at the equator, form a bowl around the formerly formed primary fibers, which thin out over time and give rise to the lens nucleus. This continuous growth of new lens fibers, without rejection of the old epithelial or fibrous cells is maintained throughout the entire life, albeit later at a much slower rate, so that the lens is constantly increasing in size and weight (Figure 3).<sup>68,72,73</sup>



### **Causes of congenital cataract**

Congenital cataract can be inherited (monogenetic or chromosomal abnormalities, metabolic diseases e.g. galactosemia) or can be caused by prenatal/postnatal infections (e.g. rubella, herpes simplex or syphilis).<sup>68</sup> Genetic factors are responsible for 8 to 25 % of affected individuals. In unilateral cataract, genetic factors are only responsible for 2 % of affected individuals, whereas in bilateral cataract gene mutations and chromosome aberrations are the main cause (in children 27 % of the cases).<sup>68</sup> Autosomal recessive and X-linked inheritance was observed in affected patients. The most frequent form of inheritance, however, is autosomal dominant.<sup>68</sup>

### **Genes implicated in cataract**

Mutations in more than 30 genes have been found to lead to congenital cataract.<sup>67</sup> Different mutations in a single gene can lead to the same phenotype, but it has been also reported that exactly the same gene mutation in two individuals can lead to different phenotypes.<sup>74</sup> Various gene mutations have been found in major cytoplasmic proteins of the human lens<sup>68</sup>:

1. Crystallins ( $\alpha$ ,  $\beta$ ,  $\gamma$  crystallin)<sup>75</sup>
  - 90 % of total lens proteins
  - Function: Important for the transparency of the lens
2. Lens specific connexin (Cx43, Cx46, and Cx50)<sup>76</sup>
  - Gap junction proteins
  - Function: Connect lens fiber cells
3. Major intrinsic protein (MIP) or Aquaporin (AQP0)<sup>77</sup>
  - Most highly expressed membrane protein in the lens
  - Function: Important for lens integrity, transport of water and small molecules
4. Cytoskeletal structural proteins (BFSP2)<sup>78</sup>
  - Function: Determination and maintenance of cell volume and shape
5. Developmental regulators (FOXE3, HSF4, MAF, PITX3)<sup>79</sup>
  - Growth factors and transcription factors
  - Function: lens plan specification

### 1.2.3 Coloboma of the eye

#### General description

Coloboma is usually caused by a failed optic fissure closure during embryogenesis (5th to 7th week of fetal life) (Figure 4).<sup>80</sup> With one affected individual out of 100,000, coloboma belongs to the rare disorders.<sup>81</sup> But even if coloboma is rare, it is responsible for 10-15 % cases of pediatric blindness.<sup>82,83</sup> A difference in prevalence between males and females was not observed.<sup>80</sup> In general, coloboma means a gap or hole in one of the major eye structures, such as the eyelids, cornea, iris, ciliary body, lens, optic nerve, retina or choroid.<sup>84</sup>

Patients can have cataract on both eyes (bilateral) or on one single eye (unilateral). Moreover, it is known that coloboma can occur isolated or in association with other syndromes. Especially, the association of coloboma with microphthalmia (small eye) or anophthalmia (absent eye) has been reported.<sup>81</sup>

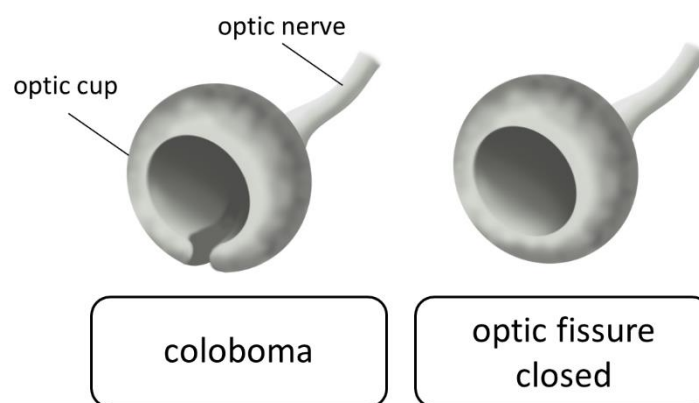


Figure 4. In patients affected by coloboma, the optic fissure did not close properly during the embryonal eye development.

#### Genes associated with coloboma

Genetic, as well as non-genetic factors can cause coloboma. Non-genetic risk factors, include e.g. an increased paternal age or an intrauterine exposure to alcohol.<sup>80,85,86</sup> It has been suggested that genetic factors, either monogenic or chromosomal, account only for 25-30 % of coloboma cases. However, many of the coloboma-associated genes have not yet been identified. The observed genetic transmission was either autosomal dominant or autosomal recessive. Interestingly, many transcriptional regulators have been associated with coloboma, such as *SOX2*<sup>87</sup>, *OTX2*<sup>88</sup>, *PAX2*<sup>89</sup>, *PAX6*<sup>90</sup>, *CHD7*<sup>91</sup>, *SALL2*<sup>92</sup>, *RAX*<sup>83</sup> and *MAF*<sup>93</sup>. Moreover, mutations have been identified in sonic hedgehog (*SHH*)<sup>94</sup> and in members of the TGF-beta signaling pathway, *GDF6*<sup>95</sup>, *GDF3*<sup>96</sup> and *MAB21L2*<sup>97</sup>. In addition, an association of members of the retinoic acid synthesis (*STRA6*<sup>98</sup> and *ALDH1A3*<sup>99</sup>) and coloboma was reported. Finally mutations in *ABCB6*<sup>100</sup>, a membrane porphyrin transporter, in *IPO13*<sup>101</sup> a nuclear transport protein and in *YAP1*<sup>102</sup>, a member

of the HIPPO growth control pathway, were identified in coloboma patients. However, there is still a large amount of unexplained cases, assuming that many further genes will be identified in the coming years.

## 1.2.4 Congenital Deafness

### **General description**

Hearing impairment or deafness is one of the most common disorders of the sensory system. The prevalence of congenital deafness varies according to the age and origin of individuals. While the prevalence in developed countries is 1 to 3.5 per 1,000 children, in developing countries about 19 children are affected by deafness among 1,000.<sup>103–106</sup> Patients with deafness are categorized in two groups according to the location of the lesion<sup>107</sup> (Figure 5A):

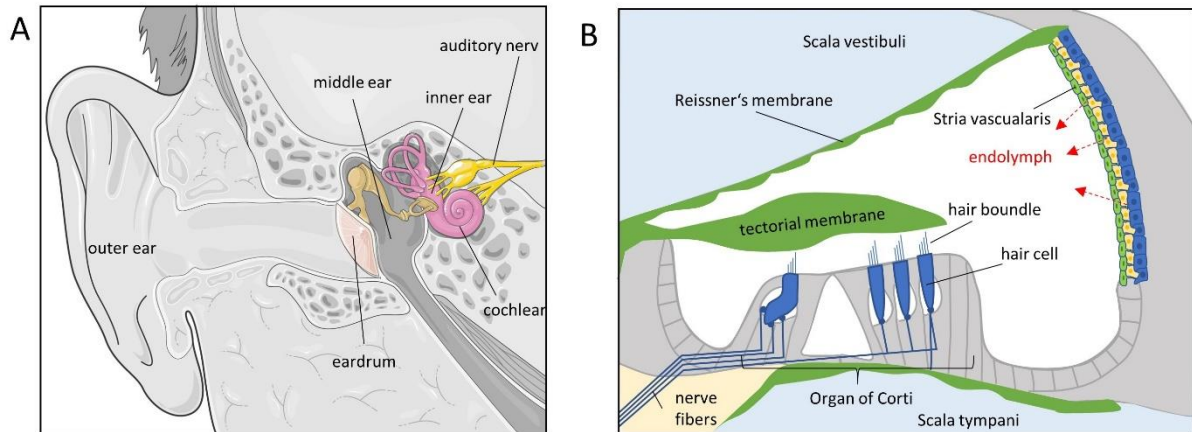
1. Conductive hearing loss: outer- and middle ear
2. Sensorineural hearing loss: inner ear, auditory nerve or central auditory pathway
  - sensory hearing loss (affected hair cells)
  - central hearing loss (central auditory pathway is affected)
  - auditory neuropathy spectrum disorder<sup>108</sup>

Responsible for congenital deafness are in most of the cases genetic factors, but also craniofacial abnormalities or congenital infections can lead to hearing impairment.<sup>107,109</sup> Gene mutations can lead to both non-syndromic and syndromic forms of hearing loss (e.g. Usher syndrome).<sup>107,110</sup> Especially, non-syndromic deafness is heterogeneous. In 80 % of non-syndromic cases, autosomal recessive deafness was reported, while an autosomal dominant inheritance was observed in the remaining 20 %. In very rare cases, maternal mitochondrial DNA-related modes were reported or X-chromosomal inherited mutations resulted in deafness.<sup>107</sup>

### **Affected genes and pathways in deafness patients**

In particular, two signal pathways are involved in hearing loss induced by genetic factors (Figure 5B). First, the correct function of the stria vascularis is essential for hearing. The stria vascularis is located on the lateral wall of the cochlear duct and is responsible for the production of the endolymph. The endolymph is the fluid in which the stereocilia are located and which ensures their mobility. For the proper function of the endolymph a precise ionic composition is necessary. An impaired ionic composition of the endolymph has been associated with hearing impairment.<sup>107</sup> Several proteins expressed in the endolymph and responsible for proper ion transport, such as GJB2, GJB6, claudin14, or MARVEL domain-containing protein 2 are associated with deafness.<sup>107,111–113</sup>

Second, malfunctions of stereocilia are suspected to be related to deafness. Mutations in genes such as *KCNQ4*, *ESPN*, *TRIOBP* or *OTOF*, which code for proteins that have an important role in stereocilia, lead to hearing loss.<sup>107,114–117</sup>



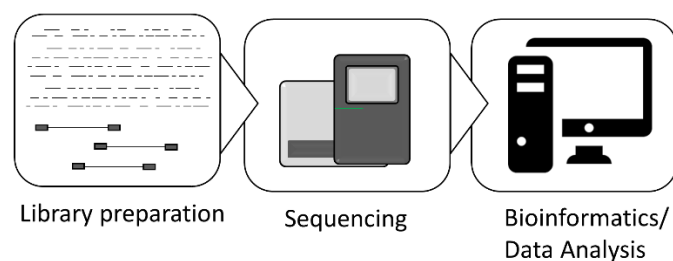
**Figure 5. Schematic representation of the human ear and cross section of the cochlear.**

**(A)** The ear is composed of an outer, middle and inner ear. (This figure was drawn using the image bank of Servier Medical Art. Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License.) **(B)** Cross section of the cochlear. The stria vascularis produces the endolymph. Against a concentration gradient, ion channels of the stria vascularis release  $K^+$  into the endolymph, resulting in a high  $K^+$  concentration. When mechanotransduction channels of the stereocilia open,  $K^+$  ions stream in, which leads to a depolarization of the hair cells and an electric signal transduction to the auditory nerve fibers. (inspired by Korver et al., 2017)<sup>107</sup>

### 1.3 Identification of novel disease genes using Next-Generation Sequencing approaches

Before the advent of NGS, causal genes for monogenic diseases were successfully identified by linkage analysis. Therefore, genome wide linkage analyses were performed in large families with many affected patients. All genes located in the identified chromosomal regions were positional candidates.<sup>118</sup> Afterwards, all positional and functional candidate genes were sequenced one by one by Sanger sequencing. The main disadvantage of this method is the fact that very large families with many affected members are required. Furthermore, Sanger sequencing can be very time-consuming, especially if many candidate genes with many exons have to be tested for mutations.<sup>119</sup> Since 2009, NGS became a powerful and popular tool in human genetics research for the identification of causative mutations in Mendelian disorders.<sup>120,121</sup> This technique allows the massively parallel sequencing of genomic regions that can be either targeted (capture or amplicon based) such as for the whole exome or non-targeted such as for the whole genome. This should provide an unbiased view of the genes of a patient. Hence, in the last years, the number of publications reporting novel disease-relevant genes increased exponentially.<sup>122</sup> Especially in the context of genetically very heterogeneous, phenotypically variable or atypical diseases, NGS technology has proven itself very efficient.

Thanks to NGS-approaches, it is possible to avoid or speedup the diagnostic odyssey of many patients with rare hereditary diseases, as the diagnosis can be made from a single wide approach.<sup>122</sup> Generally, in the interest of patients, it would be desirable to use NGS technologies earlier in molecular genetic diagnostics. Thus, a lot of time can be saved and, if available, the patient can quickly benefit from a specific therapy. In addition, the patient can thus be prevented from interventions or taking ineffective medications.<sup>123</sup>



**Figure 6. Four main steps of Next Generation Sequencing.**

There are many different types of NGS techniques, but overall, they all follow the same three basic steps, which are the creation of a library, sequencing and bioinformatics analysis (Figure 6). For the library preparation, DNA gets randomly fragmented and ligated to adapter sequences. For the sequencing of only a part of the genome, e.g. Whole Exome Sequencing (WES), a selection of only

specific sequences takes place. An enrichment of the fragments has to be performed, e.g. via bridge-amplification on a flow-cell (Illumina) and is followed by final sequencing step. The most commonly used method for sequencing is "sequencing by synthesis". The complementary integration of a fluorescently modified nucleotide during chain extension is monitored directly by the sequencer (Illumina). The light signal (fluorescence) is then registered by the machine. Subsequent data processing and filtration of variants is then performed using bioinformatics tools.<sup>124</sup>

### 1.3.1 Whole Exome Sequencing

Whole Exome Sequencing focuses exclusively on the exome, which corresponds to the protein-coding regions of the genome that is approximately 1.5 % of the entire human genome (3 billion base pairs).<sup>125</sup> In one WES experiment, approximately 40,000-80,000 variants can be identified in the 22,000 human coding genes.<sup>119</sup> WES has many advantages over traditional linkage analysis as demonstrated by Ng and colleagues. They were the first team to discover a gene mutated in a rare Mendelian disorder (Miller syndrome) by WES. They highlight that WES can be a powerful technique for gene discovery even in small families with a limited number of affected individuals or in simplex unaffected cases from multiple unrelated families.<sup>120</sup> Indeed, subsequent studies have confirmed that WES is powerful enough to identify new disease gene even in a single affected individual from a single family. In our laboratory (Strasbourg), we have also extensively used WES for gene discovery over the last decade. For example, *BBIP1* (*BBS18*) was identified as the 18th BBS gene in a single patient with Bardet-Biedl Syndrome<sup>2</sup>, a homozygous mutation was detected in *TUBGCP4* in individuals with microcephaly and chorioretinopathy<sup>4</sup>, recessive mutations in *LTBP3* were associated with brachyolmia with amelogenesis imperfecta<sup>126</sup> and a homozygous missense mutation in *VPS15* was found in three individuals with a ciliopathy phenotype.<sup>1</sup>

### 1.3.2 Whole Genome Sequencing

By contrast, Whole genome sequencing (WGS) has been used only occasionally in clinical diagnostics.<sup>127</sup> The reasons for this are the high costs for data generation and storage but this situation is bound to change soon. Above all, however, the current state of knowledge makes it difficult to assess the large number of changes in non-coding areas with regard to medical or diagnostic questions, which is why the costs and benefits of the study have to be weighed up on a case-by-case basis.<sup>128</sup> In France the "Plan France Médecine Génomique 2025" is currently addressing the use in diagnosis of WGS.<sup>129</sup>

Since WGS can analyze all ~ three billion coding and non-coding nucleotides of the human genome, WGS is also able to identify mutations in deep intronic or intergenic regulatory areas, which opens the door to a deeper understanding of mechanisms by which variation in non-coding regions contribute to

the expression of phenotypic traits.<sup>130</sup> In Table 2, the respective advantages and disadvantages of WES and WGS are listed.

**Table 2. Comparison of the advantages and disadvantages of WES and WGS.<sup>131</sup>**

	<b>Exome Sequencing</b>	<b>Whole Genome Sequencing</b>
<b>Advantages</b>	<ul style="list-style-type: none"> <li>- Identification of novel causative genes in coding regions</li> <li>- Identification of SNVs, indels, SV and CNVs</li> <li>- Low cost (compared to WGS)</li> <li>- Data are easier to interpret compared to WGS</li> <li>- Higher depth of coverage</li> </ul>	<ul style="list-style-type: none"> <li>- Identification of novel mutations in coding and non-coding regions: SNVs, indels, SV and CNVs</li> <li>- Detection of structural variants</li> <li>- Uniform coverage</li> </ul>
<b>Disadvantages</b>	<ul style="list-style-type: none"> <li>- Sequencing depth affected by poor/incomplete exome capture in biased regions</li> <li>- Cannot detect non-coding variants</li> </ul>	<ul style="list-style-type: none"> <li>- High costs</li> <li>- Large volume of data and complex analysis</li> </ul>

## 1.4 Validation of Candidate Genes using the Zebrafish model

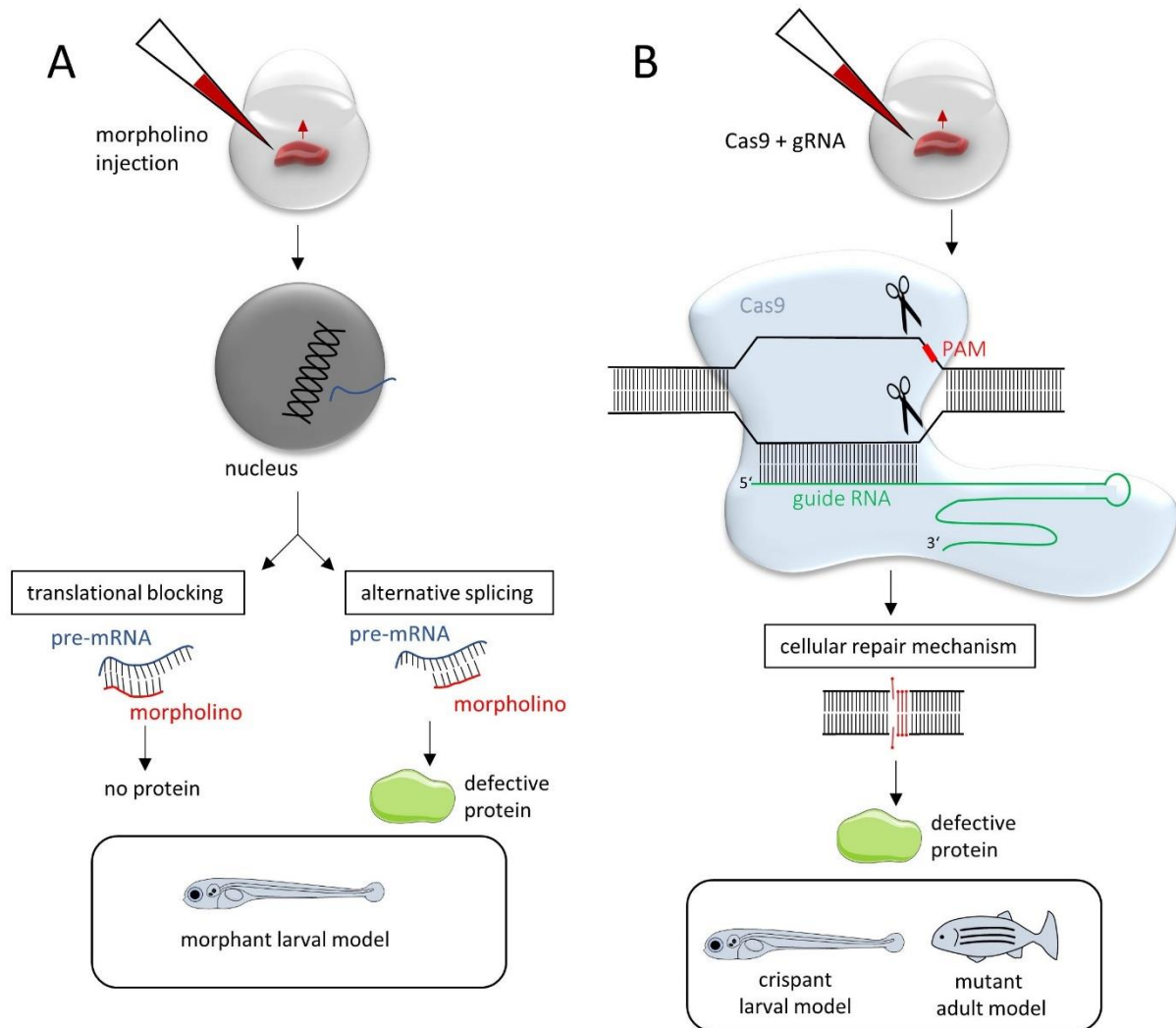
Zebrafish (*Danio rerio*) has been used as a model organism in research since 1960, but the popularity of the zebrafish model only started to increase significantly after Streisinger and his colleagues developed new techniques to facilitate their genetic analysis.<sup>132,133</sup> Although the overall morphology of fish does not resemble the human appearance, more than 70 % of all known human genes have at least one orthologue in the zebrafish genome.<sup>134</sup> The zebrafish genome is still under active development since its first release in 2013<sup>134</sup> and has been integrated the Genome Reference Consortium to maintain a high resolution reference (GRCz11 latest zebrafish assembly (<https://www.ncbi.nlm.nih.gov/grc/zebrafish>)).

Nowadays, zebrafish is a popular model organism to study various human diseases, among others also neurodevelopmental disorders and neurosensory diseases.<sup>5,135</sup>

With a full sequenced genome, a cost-effective maintenance and a high offspring number (up to 300 descendants per female per week) allowing high throughput screenings, the zebrafish model has many advantages over other animals. Moreover, embryos have a short generation time, grow externally and are transparent allowing the observation of their early development with a common light microscope. To investigate a potential link between a candidate gene and the phenotype observed in affected humans, different genetic manipulations can be performed depending on the type and transmission of a genetic disease: gene knockdown (morpholino), transgene overexpression (tagged candidate gene) or stable mutagenesis (zinc fingers, TALENs, and CRISPR/Cas9).<sup>5</sup> In Figure 7, the two methods used in this work, morpholino knockdown and CRISPR/Cas9 gene knockout, are described in more details.

Various ID/autism zebrafish models have been already investigated. To name some examples, morpholino antisense oligonucleotides have been used to knockdown the ID/autism genes *mecp2* and *kctd13*. Interestingly, in both *mecp2* and *kctd13* morphants brain anomalies were observed, such as an increased neuronal proliferation.<sup>136,137</sup> Moreover, Liu and his colleagues, generated a *shank3b* mutant, one of two orthologues of the human *SHANK3* gene, in which mutations have been associated with ASD and ID.<sup>138,139</sup> *shank3b* mutants displayed a developmental delay, a decreased locomotor activity, and an impaired social behavior.<sup>139</sup> Moreover, a zebrafish model for *SYNGAP1* (OMIM #603384), a gene associated with ASD, ID and epilepsy exists.<sup>140</sup> Zebrafish with *syngap1b* deficiency have been shown to develop brain abnormalities, such as an abnormal midbrain-hindbrain boundary, microcephaly, and increased apoptosis in the CNS. Moreover, embryos display seizure-like behaviors.<sup>141</sup> These data prove that zebrafish can be a good model to study neurodevelopmental disorders.



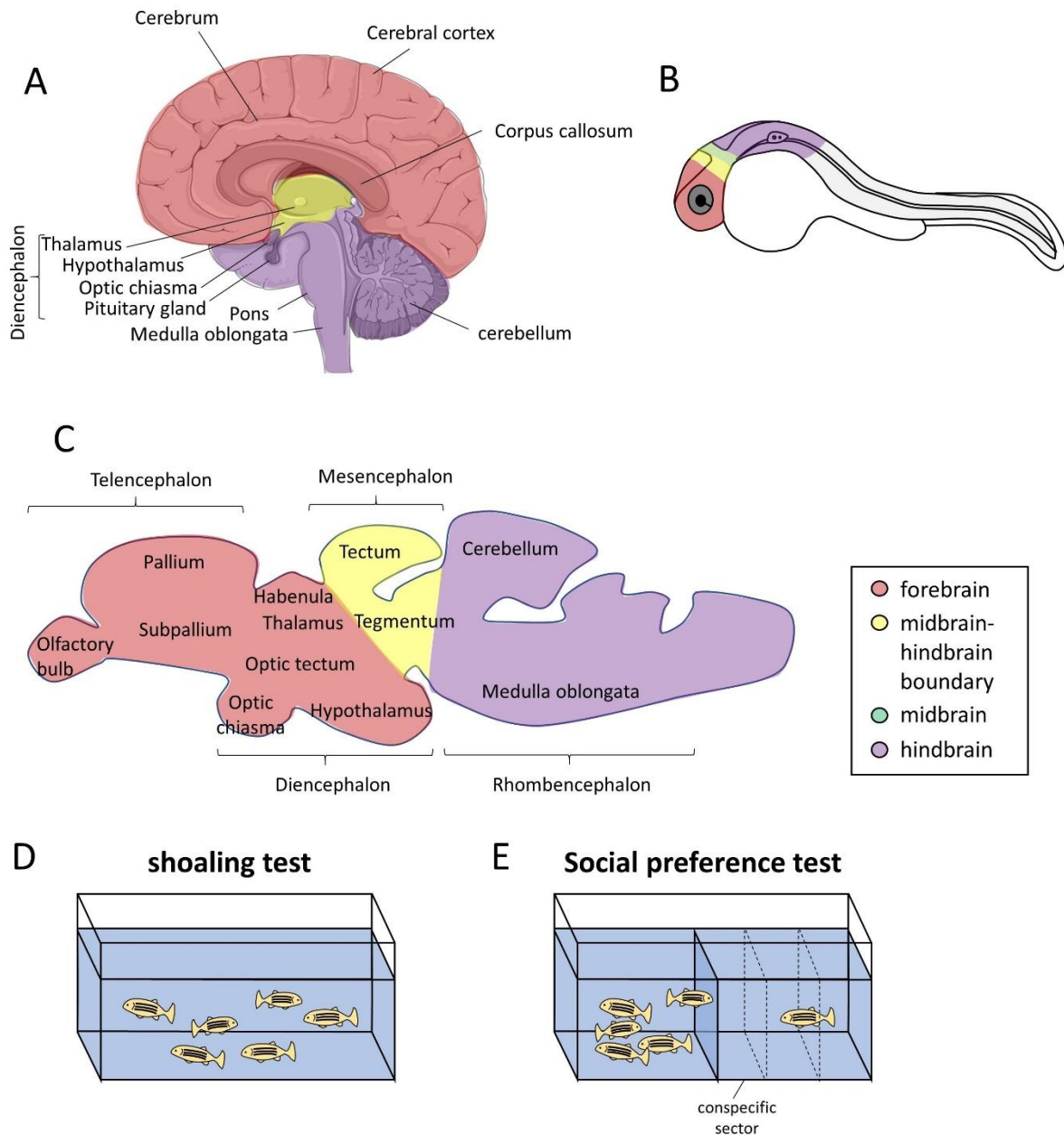


**Figure 7. Gene knockdown versus knockout in zebrafish.**

**(A)** Morpholino oligonucleotide knockdown. After the morpholino injection into one-cell stage embryos, the morpholino binds to its complementary pre-mRNA sequence and prevents its translation or affects splicing. With each cell division, the morpholino is diluted and its effect reduced, which is why the phenotype of morphants can often only be observed for up to 3 days. **(B)** Generation of knockout models using the CRISPR/Cas9 system. After the injection of a specific guide RNA and the Cas9 protein, the guide RNA directs the construct to its complementary sequence in the genomic DNA. Cas9 cleaves the DNA strand just in front of the Protospacer Adjacent Motif (PAM) sequence. Cellular repair mechanisms repair the DNA strand, which leads to small insertions and deletions of nucleotides resulting after translation in a defective protein. The created mutation is stable and is passed on from generation to generation.

#### 1.4.1 Zebrafish models for neurodevelopmental disorders

Zebrafish is a popular model for neurodevelopmental disorders, as its central nervous system (CNS) consists out of the same four main structures as in mammals: the forebrain, the midbrain, the hindbrain, and the spinal cord. Since zebrafish embryos are transparent, the development of the CNS can be easily observed as early as one day post fertilization (dpf). Moreover, many human brain structures have homologues in zebrafish, such as the thalamus, the optic tectum, or the cerebellum.



**Figure 8. Comparison between the human and zebrafish brain.**

(A) Representation of the human brain. (B) Only by 24 hpf zebrafish embryos present the three main brain structures: forebrain, midbrain and hindbrain. (C) Representation of the adult zebrafish brain. (D-E) Experiments to assess the behavior of adult zebrafish. (D) In the shoaling test fish are tested in the same tank. A camera is used to record the interfish distance. (E) In the social preference test, the preference of a single fish to swim nearby conspecifics is observed. (inspired by Vaz et al., 2019)<sup>5</sup> (The human brain was drawn using the image bank of Servier Medical Art (<http://smart.servier.com/>). Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License.)

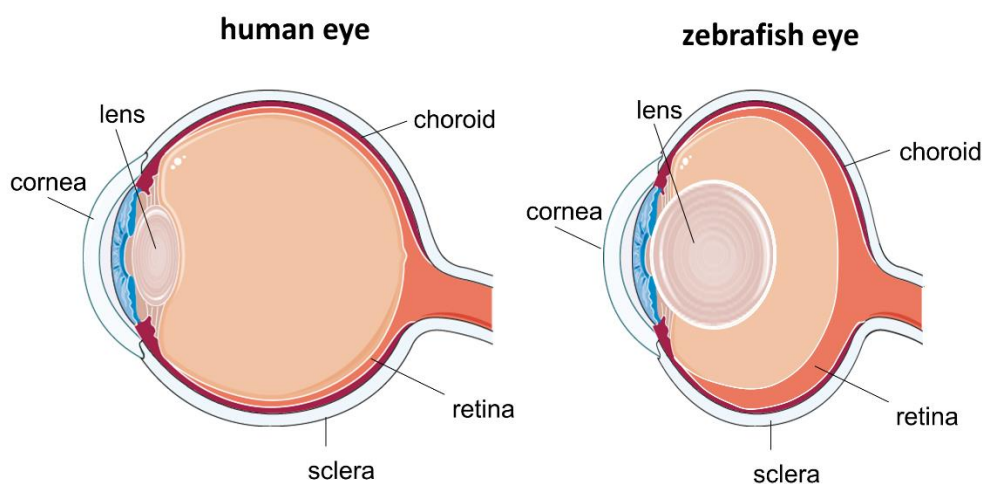
But also, clear differences can be observed between humans and zebrafish, such as the physical organization of the telencephalon and diencephalon for example (Figure 8A-C).

In addition, the most important neurotransmitter-expressing neurons, such as GABAergic, glutamatergic, monoaminergic (dopamine, noradrenaline, serotonin, histamine), cholinergic, and peptidergic neurons, are present in zebrafish.<sup>142,143</sup>

To study the impact of a candidate gene on the neurodevelopment in zebrafish, fish can be easily genetically modified. Subsequently, those genetically manipulated animals can be examined for cellular changes or behavioral abnormalities. Typical experiments to examine cellular change are e.g. *in situ* hybridizations or RT-qPCR experiments to identify mRNA expression pattern differences between wildtype fish and mutant animals. If antibodies are available, which is unfortunately not always the case for zebrafish, immunostainings can be performed to visualize protein expression alterations. Another strategy is to use transgenic fish lines to observe the brain development *in vivo* in life imaging experiments. Furthermore, an abnormal adult zebrafish behavior can be connected with an abnormal neuronal development. Already after 21 dpf, zebrafish develop a strong social behavior and preferentially form shoals.<sup>144</sup> This knowledge can be used to test behavioral anomalies. Common tests to determine the social abilities of fish are e.g. the social preference test or the shoaling test (Figure 8D-E).<sup>145</sup>

#### 1.4.2 Zebrafish models of neurosensory diseases

The overall development of the eye and inner ear as well as the underlying gene regulatory networks are highly conserved throughout evolution.<sup>70</sup> Mutations or knock-down of genes implicated in human cataract, coloboma, or deafness have been shown many times to recapitulate similar phenotypes in zebrafish<sup>146–150</sup>.



**Figure 9. Comparison of the global structure of the human and zebrafish eye.**

(The figure was drawn using the image bank of Servier Medical Art (<http://smart.servier.com/>). Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License.)

### **Zebrafish as a Model for Cataract**

The basic structures of the zebrafish eye are comparable to that of humans (Figure 9). The main difference resides in the shape of the zebrafish lens being more spherical compared to the ellipsoid human lens, resulting in a smaller vitreous chamber. Moreover, the lens vesicle is formed by delamination instead of an invagination process as seen in humans.<sup>70,151,152</sup>

There are several examples of genes which have been associated with cataract in both humans and zebrafish. For example, mutations in the human gene *HSF4* (OMIM #602438) have been identified in patients with cataract and a gene knockout in zebrafish *hsf4* causes early-onset cataract.<sup>147</sup> Moreover, *FOXE3* mutations (OMIM #601094) in human as well as in the zebrafish orthologue are known to cause cataract.<sup>151</sup> A missense mutation (P23T) in the human *CRYG* gene encoding for  $\gamma$ -crystallin has been reported to cause cataract, most likely due to protein solubility problem.<sup>153</sup> Lens defects have been also described after the overexpression of transgenic  $\gamma$ D-crystallin mutants in zebrafish.<sup>154</sup> This indicates, that zebrafish is an excellent model to study human cataract genes.

### **Zebrafish as a Model for Coloboma**

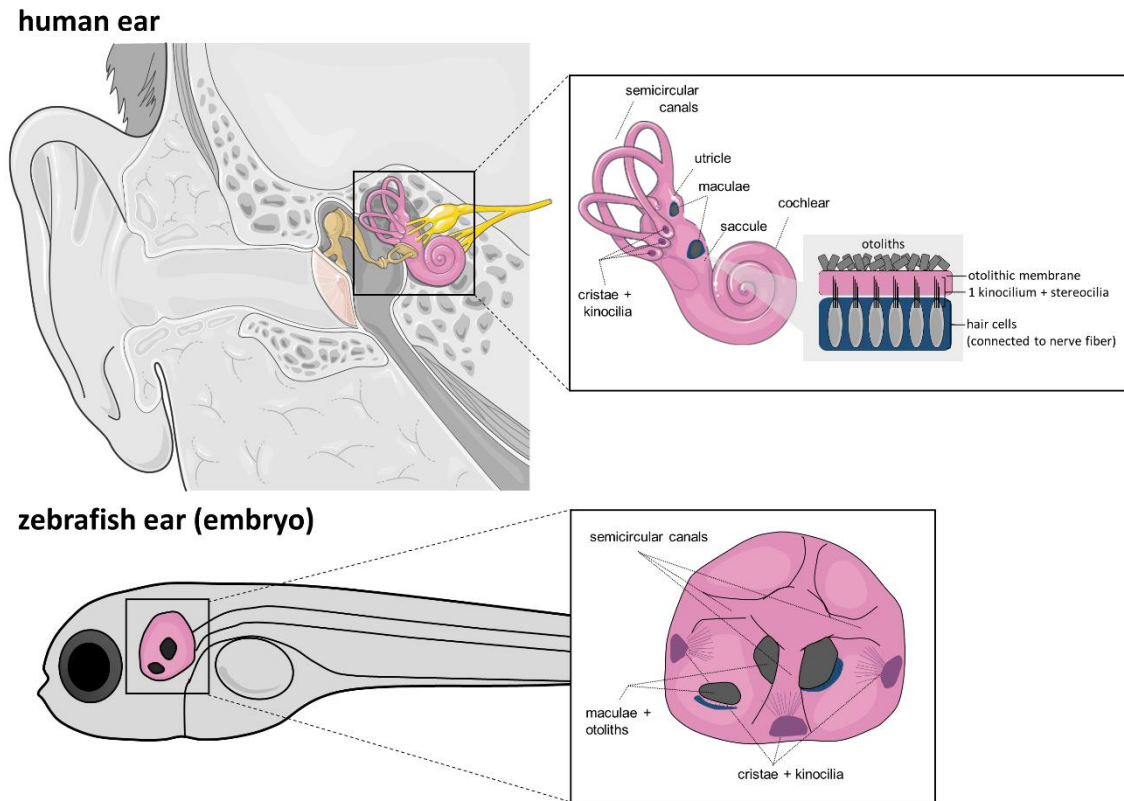
Since the optic fissure closure during embryonal eye development is highly conserved between humans and zebrafish, zebrafish have been used for various studies to get a better understanding of the pathophysiological pathways related to coloboma.<sup>155,156</sup>

For example, RNA sequencing study performed in zebrafish to examine genes relevant during optic fissure closure revealed a potential novel key player: *ntn1a*.<sup>71</sup> Moreover, zebrafish studies indicate that *nlz1* and *nlz2*, two repressors of transcription, are expressed in the optic fissure and that a morpholino induced knockdown leads to coloboma. Both genes seem to negatively regulate the expression of *pax2a* in *nlz1/nlz2* morphants.<sup>157</sup> *pax2a* is one of the human gene *PAX* orthologues, a gene associated with optic nerve coloboma as part of the Papillorenal syndrome (OMIM #120330).<sup>158</sup> In addition, various other studies in zebrafish demonstrate that *sonic hedgehog (shh)* regulates the transcription of *pax2a*. In *shh* morphants, *pax2a* is clearly reduced.<sup>156,159</sup> As coloboma has been also described in a family with a deletion in *SHH*<sup>94</sup>, the same pathways seem to be involved in the optic fissure closure of humans and zebrafish. These results highlight the value of zebrafish studies to understand the mechanisms leading to coloboma in humans. Another advantage of using zebrafish to study coloboma is, above all, the rapidity of embryonic fish development, as the optic fissure closure e.g. is terminated only after 56 hpf.<sup>71</sup>

### **Zebrafish as a Model for Hearing Impairment**

The zebrafish model has been used several times to study the balance organ or the ears of vertebrates. Unlike humans, zebrafish have no outer or middle ear (Figure 10). However, the inner ear of the fish

resembles that of other vertebrates. One of the biggest differences between the human inner ear and the zebrafish ear is probably the absence of the cochlear, a special structure responsible for human hearing. Consequently, the zebrafish model has its limitation when working on the underlying networks important for hearing.



**Figure 10. Anatomical comparison between the human inner ear and the embryonic zebrafish ear.**

Illustration of the human ear with the focus on the inner ear and the zebrafish ear after 4-5 dpf. The inner ear of zebrafish consists out of sensory maculae (associated with otoliths), cristae, three semicircular canals and is as in humans important for equilibration and hearing. Homologous structures of both species are shown in the same colors. (inspired by Pais-Roldán et al., 2016, Whitfield et al., 1996)<sup>160,161</sup> (The human ear and magnification of the inner ear was drawn using the image bank of Servier Medical Art (<http://smart.servier.com/>). Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License.)

Nonetheless, many studies have shown zebrafish to be a valuable model organism for studying hearing and deafness in vertebrates. A major advantage of zebrafish in comparison to other mammalian model organisms, is the externally development of embryos outside their mother, their transparency, and the fact, that all important structures of the ear are formed within 4 days and thus an abnormal development of the ear structures during development can be easily observed. Like humans, zebrafish also have three semicircular channels that end up in the ampoules, a thickening of the channels. Each ampoule contains a sensory crista containing a bundle of kinocilia. In addition, 4 days old embryos already have two otoliths, the utricle and saccule, both containing sensory maculae (Figure 10).

Whereas the semicircular structures are thought to play a role in equilibration, otoliths and maculae are suspected to be responsible for hearing.<sup>162</sup>

In previous publications, various zebrafish models for human deafness genes have been generated.<sup>163</sup> Mutations in the genes *MYOVIIA* and *CDH23* for example are associated with the Usher syndrome and a knockdown in the corresponding zebrafish orthologues *mariner* and *sputnik* causes deafness and splayed stereociliary bundles.<sup>164</sup> Moreover, mutations in the deafness gene *DNFA* result in sensorineural hearing loss in humans. In zebrafish, however, Busch-Nentwich and colleagues described that a morpholino targeting *dfna5* leads to disorganization of the developing semicircular canals.<sup>146</sup> To explain, why a mutation in a gene responsible for hearing in humans leads to damages of the semicircular canals is not possible yet, but this indicates that not all underlying pathways of hearing in zebrafish have been discovered yet.

II

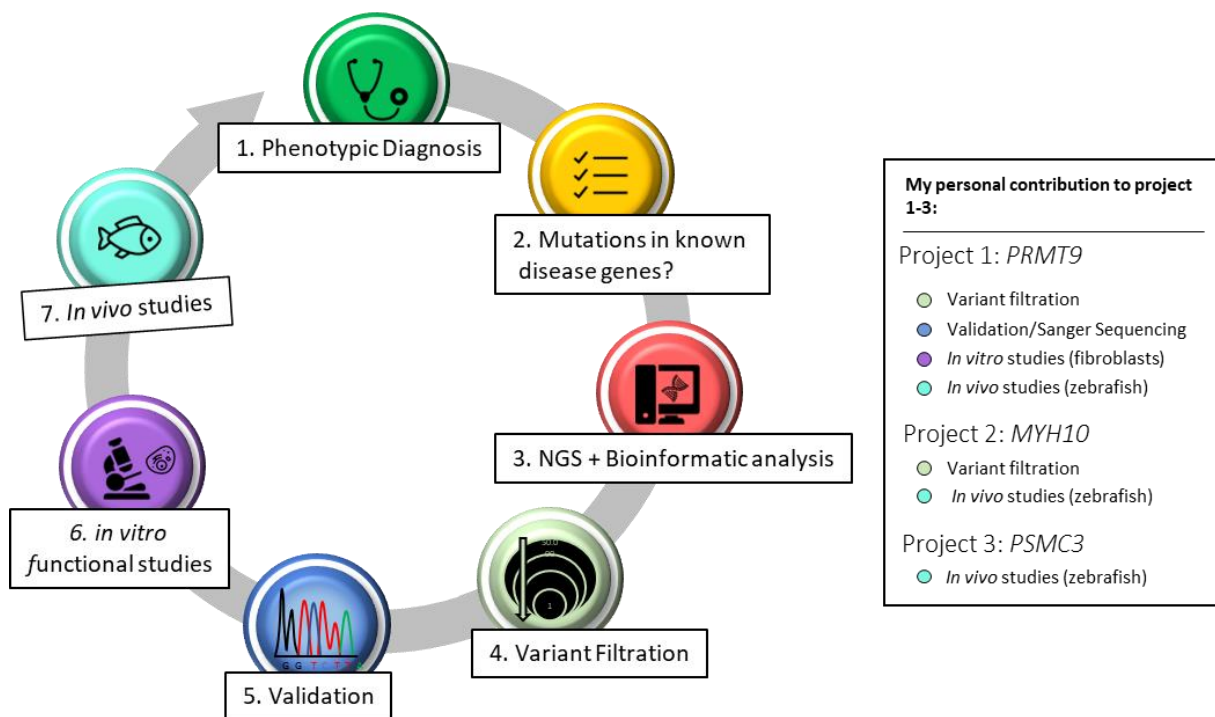
Aim of the thesis

## 2. Aim of the thesis

Neurodevelopmental and neurosensory rare disorders can dramatically affect the quality of life of affected individuals and their relatives, efficient treatment options, however, are few. Therefore, gaining new knowledge about those disorders is of great interest, especially, in order to try to identify potential therapeutic targets and pathways.

**The main aim of this thesis was, first, to perform genetic investigations on patients with neurosensory or neurodevelopmental rare diseases, in order to identify novel disease genes and second, to validate those candidate genes by functional assays to get a better understanding of the disease biology (*in vitro* and *in vivo* experiments).**

In this thesis, I report the identification of three novel disease genes associated with 3 neurosensory or neurological ultra-rare diseases. The strategy I followed, to identify those genes and my personal contribution to the three distinct projects is illustrated in Figure 11.



**Figure 11. Strategy used to identify novel disease genes and my personal contribution to the three projects of my thesis.** Icons were created with icons8.com.

For my study, I only considered patients without a genetic diagnosis and in whom previously performed Sanger sequencing and targeted next-generation sequencing (NGS) remained negative. To identify the causative gene mutation, we used NGS approaches (WES or/and WGS), followed by bioinformatics analysis, variant filtration and the validation of candidate genes *via* Sanger sequencing. Functional analyzes to validate the impact of the identified variants and the role of the gene were performed



within the laboratory of Medical Genetics (Université de Strasbourg, UMRS 1112) using patients' skin fibroblasts as well as at the Institute of Toxicology and Genetics (Karlsruher Institut für Technologie) by generating corresponding zebrafish knockdown or knockout models.

Using this strategy, I was able to contribute to the identification of three novel disease genes associated with neurodevelopmental and neurosensory disorders:

**PART 1.** *PRMT9* -> Syndromic form of intellectual disability

**PART 2.** *MYH10* -> Baraister-Winter- like syndrome (Strong ocular phenotype with coloboma and ptosis)

**PART 3.** *PSMC3* -> Neurosensory syndrome with early-onset cataract and congenital deafness

Manuscripts of all three projects have been/ are going to be submitted in the coming weeks and are included in this thesis.

**III**

**Materials and Methods**

### 3. Materials and Methods

#### 3.1 Materials

##### Bacterial Strains

All cloning steps have been performed with the competent *E.coli* strain XL1-blue from Promega.

##### Oligonucleotides

Oligonucleotides were purchased by Sigma-Aldrich or Metabion and designed using Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>).

**Table 3. Oligonucleotides**

	Name	Sequence 5' – 3'
<b>Project 1</b>		
NM_001110556.2	FLNA-ex1 F	CTGCCGCCTTCTCCAAC
	FLNA-ex1 R	GTGCGGCCAACAAAGAAGG
	FLNA-ex31/32 F	ACTAGGGGCTCACAGAACAC
	FLNA-ex31/32 R	AACAGACTCTCCAGCAGCTC
NM_001289745.1	GAPDH F	TGATGACATCAAGAAGGTGGTGA
	GAPDH R	TCCTTGGAGGCCATGTGGGCCAT
NM_138364.4	PRMT9-RT-ex3 F	CACTTGGATTGGCAGGATTT
	PRMT9-RT-ex3 R	TCCCAAATTTTCTTAGCAGT
	PRMT9-RT-ex9 F	CAAAGGCCAAAATCAGACAA
	PRMT9-RT-ex10 R	ATTCCACAAGCAACCCAAAC
NM_152891.2	PRSS33-RT-ex4 F	GCACTGCCAGCTGAGTACC
	PRSS33-RT-ex4 R	GTAGCGGTCGCCACTCTG
NM_001165963.3	SCN1A-ex15 F	GGCCAGGAAGCATGAAGGAT
	SCN1A-ex15 R	CCCATTGTTACCTGGGCTCT
NM_001101.5	β- Actin F	TGGCACCCAGCACAAATGAA
	β- Actin R	CTAAGTCATAGTCCGCCTAGAAGCA
NM_032961.3	PCDH10 F	GTCCAACGAGACTAAACACCAG
	PCDH10 R	TTCCTCAGTGCAATTGGAGA
NM_032812.9	PLXDC2 F	TGAAATCACTGTGGCAACCG
	PLXDC2 R	CAAGTGCTGTGCCATTATCAA
NM_020717.3	SHROOM4 F	AGCCGCTCACAGTGTCTAAG
	SHROOM4 R	GCGTTCCTCCTCTGACAAT
NM_022337.3	RAB38 F	GGGATATCGCAGGTCAAGAA
	RAB38 R	CCCCTGGTCACATTTGTTGG
NM_005335.6	HCLS1 F	GAAAAGGAGCAACGATGGGG
	HCLS1 R	GCACTCTGTCCATTCCGGTC
NM_001371271.1	GLI2 F	AAGCAAGAAGCCAAAAGTGG
	GLI2 R	TGGTACCTTCCTCCTGGTG
NM_001165963.3	SCN1A F	GAGGACCTGGACCCCTACTA
	SCN1A R	ACAGTTTGTCAAATAGTGACA
NM_015272.5	RPGRIP1L F	GCAAGCTACAGATCAAAGGTCA
	RPGRIP1L R	GTGGCTGATTCTGAGAGTTCT
NM_001110556.2	FLNA F	TCCATCGACAGCAAGGCC
	FLNA R	CTGAAGTTGGTGATGGGCAG
NM_000434.4	NEU1 F	CCAAGTTCATCGCCCTGC
	NEU1 R	TCCTTGCTCCATACCAACATG

Table 3. (continued)

RefSeq/ Ensemble	Name	Sequence 5' – 3'
<b>Project 1 (continued)</b>		
NM_004335.4	BST2 F	CACCTGCAACCACACTGTG
	BST2 R	TCTCTTCTCAGTCGCTCCAC
NM_002612.4	PDK4 F	CAGTGGTCCAAGATGCCTTT
	PDK4 R	GCATATGATGGAGGTGAGAAGG
NM_139314.3	ANGPTL4 F	CCACCGACCTCCCGTTAG
	ANGPTL4 R	CTTCTCCAGGTGCCGCTG
NM_000168.6	GLI3 F	GGCATTTTTGGTCGAAGAGA
	GLI3 R	GGACATTCTGTGGCTGCATA
NM_001256799.3	GAPDH-RT-F	GGAGCGAGATCCCTCCAAAAT
	GAPDH-RT-R	GGCTGTTGTCATACTTCTCATGG
NM_000194.3	HPRT1-RT-F	CCTGGCGTCGTGATTAGTGAT
	HPRT1-RT-R	AGACGTTTCAGTCCTGTCCATAA
Zebrafish ENSDARG00000036755	prmt9-ex4 F	GAGGTCTATGCCTGTGAGCTGT
	prmt9-ex4/int R	TTTCAACAAGACTTTTATTACCTG
	prmt9-ex8 F	AGGCCATCTATCCAGTGCAG
	prmt9-ex9 R	GCAACACAGCATAGCCTCAG
Zebrafish ENSDARG00000037746	bactin F	GCCTGACGGACAGGTCAT
	bactin R	ACCGCAAGATTCCATACCC
<b>Project 2</b>		
NM_005964.3	MYH10-ex32F	TTTTTCGTATCCCCATCC
	MYH10-ex32R	TCCTAGATGGCTTTGCTGCT
	MYH10-RT-ex30F	AGGAGGCCAGGAAGAACCT
	MYH10-RT-ex33R	CTCCAGCTGGGTCCTCATT
	MYH10-RT-ex2F	GGCAGAATTGACATGCTTGA
	MYH10-RT-ex3R	AGGCATCTCATGACGCTTCT
	MYH10-RT-ex35F	AAGCTCCAGGCTCAGATGAA
	MYH10-RT-ex36R	GCTCAGATGAGGCAAGTTCC
NM_001289745.1	GAPDH-RT-ex4-5F	GGAGCGAGATCCCTCCAAAAT
	GAPDH-RT-ex6R	GGCTGTTGTCATACTTCTCATGG
NM_004850.4	ROCK2-ex27F	CAGGCAGTAAGAATACTTTTTGGA
	ROCK2-ex27R	TGTTACTTGAATCCTTACTCCAAGG
NM_006289.3	TLN1-ex29F	CTGCATCATTCCCCACTCTT
	TLN1-ex29R	AGTGGAACCATTC AACCAG
NM_001277224.1	TAGLN2-int2F	GGTTGAGGGAGGGTAGATCC
	TAGLN2-int2R	CTTGACAAAGGCCTTCAAGC
NM_001436.3	FBL-ex4F	GGCGGAGATTCAGGATCTTT
	FBL-ex4R	GGGAGCAGACAGTGTGGTTT
Zebrafish ENSDARG00000000103	myh10-5'UTRF	cttgaattcgtggcctatttttagttcctcATGCCGGAGATG
	myh10-ex1R	CTTCTCGAGATATGAGTCCAGAGTAATAACGATCCTTCAGGT
	myh10-ex3F	GGACTCTTCTGCGTGGTCAT
	myh10-int5-6R	AACTCTGAGGACAGCCGAGA
<b>Project 3</b>		
NM_001205119.1	ATG13-RT-ex9F	CAGGAAACAAGGGCATGAAT
	ATG13-RT-ex14R-15R	AGGCGAGAGCTTGAGAG7TG
NM_002804.4	PSMC3-ex10F	CTCAGGTAATTGCAGCCACA
	PSMC3-int10R	CGCCTGTAGTCCCAGCTACT
	PSMC3-int10R	GCGACAGAGCGAGACTG
	PSMC3-RT-ex9-10F	AAGTTAAGGTAATTGCAGCCACA
	PSMC3-RT-ex12R	CATGTAGTCTCGTGGGTGA
	PSMC3-RT-int10F	ACAGAGGCTGGAGGCACTTA
	PSMC3-RT-int10R1	TTCAGATGAGGGGTGGAGTC
PSMC3-RT-ex11R	GCCAGCTCCTCGTAGTTCAC	

Table 3. (continued)

	Name	Sequence 5' – 3'
<b>Project 3 (continued)</b>		
Human <i>PSMC3</i> REF RefSeq Gene NM_002804.4	PSMC3-RT-int10-ex11R	TCACGTCAGGACTGAACTCAA
	PSMC3-RT-QPCR-ex5-6F	TAAAGCCAGGAGACCTGGTGGGTG
	PSMC3-RT-QPCR-ex7R	GCCCATACATCAGCACCCCTTT
	PSMC3-RT-QPCR-ex11F	GACGTGAACTACGAGGAGCTG
	PSMC3-RT-QPCR-ex12Rbis	GCACTTCAGCCGTGAGACT
Zebrafish <i>psmc3</i> ZDB-GENE-030131-666 (ENSDART00000171691.2)	psmc3-inSitu-ex2F	GAAGAGATTGTTTCAGAGGACTCG
	psmc3-inSitu-ex9R	GGGCATAGGGAAGTCAATCTTA
	psmc3-full_length-ex1F	GCAGAATTCATGTCGTCGCTGAATGACAGA
	psmc3-full_length-ex11R	TTAGCGGCCGCTTAAGCATAGTACTGCAAATT
	psmc3-mo-RT-ex2F	GAAGAGATTGTTTCAGAGGACTCG
	psmc3-mo-RT-ex4R	TGCCGTGTGGAGTTTTGAT

Table 4. Oligos ordered for the synthesis of CRISPR guide RNAs and the corresponding primers used for genotyping.

Gene	Oligos for guide RNA synthesis	Binding sequence (PAM)	Primers (Genotyping)
<b>Project 1</b>			
<i>prmt9</i> -gRNA	taatacagactcactata <u>GGTCCAGAG</u> ATTTCCGGTGAgtttagagctagaa	TCTCCAGAGATTTCCGGTGA <u>AGG</u>	<i>prmt9</i> .ex5F GAGGTCTATGCCTGTGAGCTGT <i>prmt9</i> .exR TTTCAACAAGACTTTTATTACCTG
<b>Project 2</b>			
<i>myh10</i> -gRNA	taatacagactcactata <u>GTTCTCAGAG</u> TATATGGGGgtttagagctagaa	TGTTCTCAGAGTATATGGGG <u>AGG</u>	<i>myh10</i> .ex3F CATACTCAGGACTCTTCTGCGTG <i>myh10</i> .ex3R GCATGCATCTGTATGCCG
<b>Project 3</b>			
<i>psmc3</i> -gRNA1	taatacagactcactata <u>GGATCCTAAT</u> GACCAAGAGGgtttagagctagaa	AGATCCTAATGACCAAGAGG <u>AGG</u>	<i>psmc3</i> .in3-4F GGCTGGAGTATGGTGTTAAAGC <i>psmc3</i> .int4-5R AAAGATGGATGGAAGAATTTGG
<i>psmc3</i> -gRNA2	taatacagactcactata <u>AGGATATCC</u> ACCCTGTTAGgtttagagctagaa	CAGGATATCCACCCTGTTAGT <u>GG</u>	<i>psmc3</i> .ex8F TGGAGCTGCTCAATCAGTTAGA <i>psmc3</i> .ex8R GGGCATAGGGAAGTCAATCTTA

## Plasmids

Table 5. Plasmids used in this study.

name	reference
pCS2+	NovoPro
pCS2+gfp	
pGEM-T easy	Promega

**In situ hybridization probes****Table 6. In situ hybridization probes used in this study.**

name	reference
<b>Project 1</b>	
<i>prmt9</i>	this study (probe binds to 552 bp of the <i>prmt9</i> transcript)
<b>Project 2</b>	
<i>pax2a</i>	Armant <i>et al.</i> (2013) <sup>165</sup>
<b>Project 3</b>	
<i>psmc3</i>	this study (probe binds to 956 bp of the <i>psmc3</i> transcript)
<i>her8a</i>	Armant <i>et al.</i> (2013) <sup>165</sup>
<i>krox20</i>	
<i>msxc</i>	
<i>sox19b</i>	
<i>otopetrin</i> <i>versican a</i> <i>versican b</i>	provided by Tanya Whitfield (University of Sheffield, United Kingdom)

**Zebrafish lines****Table 7. Zebrafish lines used in this study.**

Transgenic/ mutant	allele name	description	reference
<b>Project 1</b>			
<i>prmt9</i> +4bp	ka709	ENSDART00000053369.6: c.707_708insAAAT; p.His236Glnfs*33	this study
<b>Project 2</b>			
<i>myh10</i> -7bp	ka708	ENSDART00000151114.2: c.405_411delCATATAC; p.Ile135Leufs*97	this study
<b>Project 3</b>			
gSALzGFFM539A	-	ENSDARG00000006350: Chr. 13 41385444-41385002, insertion into exon of <i>nkx1.2la</i> (expression in spinal cord neurons and floorplate, canals of the ear from 2 dpf onwards)	generated using the Gal4-UAS system as described previously <sup>166</sup> by Koichi Kawakami (Laboratory of Molecular and Developmental Biology, National Institute of Genetics, and Department of Genetics, SOKENDAI (The Graduate University for Advanced Studies), Mishima, Japan)

## Morpholinos

**Table 8. Morpholinos used in this study.** Morpholinos were all purchased by Genetools LLC (Oregon).

Name	Sequence (5' – 3')
<b>Project 1</b>	
<i>prmt9-mo</i>	GTCTTGTAGTGGCATTAGGCATCTC
<i>ctrl-mo</i>	GTGTTCTACTGCCATTACGCATCTC
<b>Project 2</b>	
<i>myh10-mo</i>	GCCATCTCCGGCATGAGGAACTAAA
<i>ctrl-mo</i>	GCgATgTCCcGCATcAGcAACTAAA
<i>myh10-mo2</i>	GCCGAGAGAAAAGAACTCACTGGA
<i>ctrl-mo2</i>	GCCcAcAGAAAAcAAACTCAgTcGA
<b>Project 3</b>	
<i>psmc3-mo</i>	TGTGAATCACAGTATGAAGCGTGCC
<i>ctrl-mo</i>	TGTcAATgAgAGTATcAAcCGTGCC

## Antibodies

**Table 9. Antibodies used in this study.**

Name	Producer	Application(s)	Conditions
<b>Project 1</b>			
Mouse anti-acetylated-Tubulin	Invitrogen (Cat. #32-2700)	IF	1/400 ; 5 % BSA
Mouse monoclonal anti-PRMT9	Merck Millipore (Cat. #128-29-1)	WB	WB: 1/500; 5 % milk
Mouse monoclonal anti- $\beta$ -Tubulin	Euromedex (Cat. #322700)	WB	1/2500 ; 5 % milk
Rabbit polyclonal anti-SAP145	Novus Biologicals (Cat. #NB100-79848)	WB	1:3000 ; 5 % milk
Rabbit anti-SDMA (BL8244)	Yang et al., 2015 <sup>167</sup>	WB	1:1000 ; 5 % milk
Goat anti-mouse IgG HRP	Santa Cruz Biotechnology (Cat. #NC9340011)	WB	1:5000 ; 5 % milk
Goat anti-mouse Alexa Fluor 594	Invitrogen Cat. # A-11005	IF	1:500 (in 5 % BSA)
Donkey anti-Rabbit HRP ECL Rabbit IgG, HRP-linked	GE Healthcare Life Sciences Product code: NA934V	WB	1:5000 ; 5 % milk
<b>Project 2</b>			
Rabbit Anti-MYH10	NOVUS NBP2-38566 (BioTechne)	Western, IF	WB: 1/1000 milk 5 %, IF: 1/100
Mouse Anti-myosin (A.1025)	Developmental Studies Hybridoma Bank	IF	IF: 1/100
Rabbit Anti-GAPDH	Abcam #ab181602	Western	WB:1/5000c
Rabbit Anti-ARP3	Abcam #181164	IF	IF:1/200
Donkey anti-Rabbit HRP ECL Rabbit IgG, HRP-linked	GE Healthcare Life Sciences Product code: NA934V	Western	WB: 1/5000; 5 % milk
Donkey anti-Rabbit IgG (H+L) Alexa Fluor 488	Thermo Scientific #A21206	IF	IF: 1/1500

Table 9. (continued)

Name	Producer	Application(s)	Conditions
<b>Project 3</b>			
Rabbit anti-PSMC3	Novus Biologicals NBP1-86962	WB. IF	WB: 1/1000; 5 % milk IF: 1/500
Rabbit anti-PSMC3	Abcam ab171969	WB. IF. IP. MS	WB: 1/2000; 5 % milk IP/MS: 10µl/mg proteins IF: 1/100
Rabbit anti-ACTG1	Novus Biologicals NB600-533SS	WB	WB: 1/10 000; 5 % milk
Rabbit anti-GAPDH	Abcam #ab181602	WB	WB: 1/5000; 5 % milk
Mouse anti-ubiquitin	Invitrogen #13-1600	WB	WB: 1/1000
Mouse Anti-acetylated Tubulin	Sigma-Aldrich T7451	IF	IF:1/500
Goat anti-Rabbit Alexa Fluor 568	ThermoFisher Scientific A-11011	IF	IF: 1/750
Donkey anti-Rabbit HRP ECL Rabbit IgG. HRP-linked	GE Healthcare Life Sciences NA934V	WB	WB: 1/5000; 5 % milk
Mouse anti-Rabbit	Cell Signalling #5127	WB	WB: 1/2000; 5 % milk
IF: immunofluorescence. IP: immunoprecipitation. MS: mass spectrometry. WB: Western Blot.			

## Solutions and buffers

**Table 10. Solutions and buffers used in this study.** If not specified differently, all solutions were prepared with deionized water.

Name	Recipe (final concentration)
20x SSC-Buffer	3 M NaCl, 0.3 M sodium citrate
Blocking buffer	1x PBS, 0.1 % v/v Tween20, 0.2 % w/v BSA, 1 % v/v DMSO
BT-Fix	4 % w/v PFA, 4 % w/v sucrose, 0.12 mM CaCl <sub>2</sub> , 0.2 M Na <sub>2</sub> HPO <sub>4</sub> , 0.2 M NaH <sub>2</sub> PO <sub>4</sub>
E3 medium (60x)	5 mM NaCl, 0.17 mM KCl, 10 mM HEPES, 0.33 mM MgSO <sub>4</sub> , 0.33 mM CaCl <sub>2</sub>
Hybridization (HYB) buffer	50 % v/v formamide, 5x SSC, 500 µg/ml yeast RNA, 50 µg/ml heparin, 0.1 % v/v Tween20, 9 mM citric acid
LB-medium	1 % w/v Bacto-Trypton, 0.5 % w/v yeast extract, 1 % w/v NaCl
Proteinase K solution	10 µg/ml proteinase K in PTW
PTW	1x PBS, 0.1 % v/v Tween20
PTU solution	0.0003 % w/v phenylthiourea, 1x E3 medium
Staining buffer	100 mM Tris-HCl pH 9.5, 50 mM MgCl <sub>2</sub> , 100 mM NaCl, 0.5 % v/v Tween20
Wash buffer 1	50 % formamide, 50 % 2x SSC; 0.1 % v/v Tween20
Wash buffer 2	2x SSC, 0.1 % v/v Tween20
Wash buffer 3	0.2 % SSC, 0.1 % v/v Tween20
TAE (50x)	2 M Tris base, 1 M acetic acid, 50 mM EDTA (pH8)
RIPA buffer	50 mM Tris pH 7.5, 150 mM NaCl, 2 mM EDTA, 1 mM N-ethylmaleimide, 10 µM MG-132, 1 % NP40, 0.1 % SDS
Lysis buffer (WB)	20 mM Tris HCl (pH 8), 137 mM NaCl, 10 % Glycerol, 1 % Nonidet P-40 (NP-40), 2mM EDTA



**Software****Table 11. Software used in this study.**

Name	Used for...	Source
ApE	Illustration of Sanger Sequencing results	Wayne Davis from the University of Utah
ChopChop	Design of guide RNAs	<a href="http://chopchop.cbu.uib.no/">http://chopchop.cbu.uib.no/</a>
GraphPad Prism 6/8	Statistics	<a href="https://www.graphpad.com/scientific-software/prism/">https://www.graphpad.com/scientific-software/prism/</a>
Ligation calculator	Ligation	<a href="http://www.insilico.uni-duesseldorf.de/Lig_Input.html">http://www.insilico.uni-duesseldorf.de/Lig_Input.html</a>
Microsoft Excel	Data analysis	Microsoft office
Microsoft Power Point	Images	Microsoft office
Primer3 Plus	Primer design	<a href="http://primer3.ut.ee/">http://primer3.ut.ee/</a>

## 3.2 Methods carried out in the Laboratory of Medical Genetics

The Laboratory of Medical Genetics has a vast experience in the identification of novel rare disease genes. The team consists out of professionals of various fields, including medicine, bioinformatics and biology. Thanks to this teamwork, I had the opportunity during my PhD to be part of the identification of three novel disease genes. Human genetic clinical diagnosis, blood sampling and skin biopsies were performed by Sophie Scheidecker, Elise Schaefer and H el ene Dollfus as well as by coworkers of the reference center for genetic ophthalmologic diseases (CARGO) or the department of medical genetics of the Strasbourg University Hospital. In order to identify the pathogenic mutation(s) responsible for the disease of patients presenting with a phenotype compatible with an inherited disorder, the genome or the exome of the patients was sequenced. The bioinformatic analysis of this complex data was done by V eronique Geoffroy and Jean Muller. The final filtration of the gene variants, the cultivation of patients' cells as well as molecular and biochemistry methods were carried out by myself, Corinne Stoetzel and Aline Schneider.

### 3.2.1 Patient Selection

The laboratory of medical genetics UMRS\_1112 works in close collaboration with the service of medical genetics and the CARGO at the Strasbourg University hospital allowing them a direct contact with patients and ad hoc recruitment for research. Following our gene identification, further families were recruited using the online matching tools Decipher<sup>168</sup> and GeneMatcher<sup>169</sup>.

**Project 1 | Family A.** Due to the presence of bilateral postaxial polydactyly and developmental delay, the index patient was referred to the Genetic Diagnosis Laboratory at the Strasbourg University Hospital with a suspicion of BBS. The absence of pathogenic mutations in known BBS genes led to the inclusion of this patient in a research project to identify novel genes mutated in BBS. The parents of the patient were informed about the objectives of the study and their consent was obtained according to the French legislation. This study was approved by the local ethics committee at the Strasbourg University Hospital.

Family B. This family was already published by Najmabadi et al.<sup>170</sup>

Family C-I. Using the online matching tools Decipher<sup>168</sup> and GeneMatcher<sup>169</sup>, we were able to recruit seven further families with syndromic ID. Study protocols used in each cohort have been approved by the corresponding Institutional Review Board or equivalent committees (in Strasbourg, "Comit e Protection des Personnes" EST IV, N oDC-20142222), and written informed consent was given by each participant or parents. Our research complies with the Declaration of Helsinki. Written informed consent for open-access publication was provided by the participants or their parents.

Family A	Pr H. Dollfus/Dr Roume
Family B	Dr Kahrizi/Dr Najmabadi
Family C	Dr A. van Haeringen
Family D	Pr F. Hildebrandt/Dr Kolvenbach
Family E	Dr B. Keren/Dr S Whalen
Family F	Dr F. Laconne/Dr S. Kircher
Family G	Dr Kristian Tveten/Dr M. Smeland
Family H	Dr K. Retterer/Dr K. McWalter/Dr Ebba Alkhunaizi/ Dr. Chumei Li
Family E	Dr K. Retterer/Dr K. McWalter/Dr Roman Yusupov

**Project 2** | The family was referred to the CARGO (*Centre de référence pour les Affections Rares en Génétique Ophtalmologique*, a French reference center for rare eye diseases at the Strasbourg University Hospital, France) because of unusual ophthalmic presentation and is followed by Hélène Dollfus. The research we carried on the patients followed the tenets of the Declaration of Helsinki. Approval was obtained from the local research ethics committee by the Strasbourg University Hospital ethics committee as part of the ultra-rare disease cohort. Informed consent was obtained as an essential prerequisite for study inclusion. The proband (individual II.2) benefited from a skin biopsy for cell biology assays (following a protocol described elsewhere).<sup>171</sup> This research followed the tenets of the Declaration of Helsinki. Approval was obtained from our institutional review board “*Comité Protection des Personnes*” (EST IV, N°DC-20142222). The identified gene was submitted to the GeneMatcher tool but no other patient with the same phenotype could be identified.<sup>172</sup>

**Project 3** | The 3 cases (II.2, II.4, and II.7) have consulted independently and were enrolled subsequently by the CARGO (reference center for rare eye diseases at the Strasbourg University Hospital, France) and are followed by Hélène Dollfus. All participants were assessed by a clinical geneticist, a neuro pediatrician, an ENT specialist, a dermatologist and a pediatric ophthalmologist. Written consent for research and publication was obtained for all study participants. This research followed the tenets of the Declaration of Helsinki. Approval was obtained from our institutional review board “*Comité Protection des Personnes*” (EST IV, N°DC-20142222). The identified gene was submitted to the GeneMatcher tool but no other patient with the same phenotype could be identified.<sup>172</sup>

### 3.2.2 Genetic analysis

#### **Homozygosity mapping**

In case of recessive inherited disorders and a suspected consanguinity, homozygosity mapping is a valuable tool to identify large regions of homozygosity within a patient. In combination with linkage analysis it has been used many times in the past to detect recessive gene mutations in consanguineous families. In the last decade, there has been a huge revolution of technologies and today the

identification of novel disease genes mainly relies on WES or WGS experiments.<sup>173</sup> As **project 1** (2008) and **3** (2010) started already about 10 years ago, homozygosity mapping was performed in both cases in the very beginning of the study. Today, the technique is mainly used to reduce the immense number of variants detected by WES or WGS. The Affymetrix GeneChip® Mapping 250K Array Xba 240 (**project 1**) or GeneChip® Mapping 350K Array Xba 240 (**project 3**) (Affymetrix, Santa Clara, CA) were used to identify homozygous SNPs in the probands. Sample processing and labelling were performed according to the manufacturer's instructions. Arrays were hybridized on a GeneChip Hybridization Oven 640. Then they were washed with the GeneChip Fluidics Station 450 and scanned with a GeneChip Scanner 3000. To generate SNP allele calls, the data were processed by the GeneChip DNA Analysis Software version 3.0.2 (GDAS). An average call rate >99 % was obtained. We only considered regions of homozygosity longer than 25 adjacent SNPs. The HomoSNP software (IGBMC, Strasbourg) was used to visualize shared regions of homozygosity.

**Project 1 |** In project 1, autozygosity mapping was performed in the index patient (individual A. II-1) to discover large regions of homozygosity that were identical by descent. The candidate gene *PRMT9* lies within a large homozygosity region on chromosome 4 (supplemental Table S1.1).

**Project 3 |** Three affected individuals (II.2, II.4, and II.7) and 2 unaffected individuals (II.1 and II.3) were analyzed. Three regions of homozygosity of respectively 0.3, 2.1 and 1.3 Mb on chromosome 11 (11:44,396,024-44,668,374; 11:45,574,574-47,684,908; 11:66,066,993-67,349,899) are shared between the affected and not the healthy individuals (supplemental Figure S3.1).

### **Whole-Exome Sequencing**

Whole exome sequencing was performed by Integragen, Evry, France (**project 1**, individual A.II-1) or by the IGBMC (*Institut de Génétique et de Biologie Moléculaire et Cellulaire*, Illkirch-Graffenstaden, France) Microarray and Sequencing platform (**project 2 and 3**). Sequencing libraries were prepared starting with 1 µg of genomic DNA. DNA was fragmented, and exons were enriched with the in-solution SureSelect Target Enrichment System (Human All Exon 50 Mb kit, Agilent, USA). Paired end high-throughput sequencing was performed with the Illumina HiSeq 2000 analyzer yielding 75 bp paired-end reads (**project 1**) and the Illumina HiSeq 2500 instrument to generate 100 bp paired-end reads (**project 2 and 3**). In order to detect gene variants, the reads were aligned to the human reference genome GRCh37/hg19 using the Burrow-Wheeler aligner (BWA v0.7.5a/ v.0.7.1).<sup>174</sup> Indel realignment, base quality score recalibration and variant calling of single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels) was performed using the Genome Analysis Toolkit (GATK v3.1/ v3.4-46).<sup>175</sup> Structural Variations (SV) were called using CANOES.<sup>176</sup> Annotation and ranking of SNV and indel were performed by VaRank<sup>177</sup> in combination with the Alamut Batch software (Interactive Biosoftware, Rouen, France). Variant effect on the nearest splice site was predicted using

MaxEntScan<sup>178</sup>, NNSplice<sup>179</sup> and Splice Site Finder<sup>180</sup>. Very stringent filtering criteria were applied to filter out non-pathogenic variants: (i) variants represented with an allele frequency of more than 1 % in public variation databases including the 1000 Genomes<sup>181</sup>, the gnomAD database<sup>182</sup>, the DGV<sup>183</sup> or our internal exome database, (ii) variants in 5' and 3' UTR, downstream, upstream, intronic and synonymous locations without pathogenic prediction of local splice effect. Annotation of SV were performed by AnnotSV<sup>184</sup>. Variations have been classified according to the ACMG guidelines.<sup>185</sup> The Integrative Genomics Viewer (IGV) was a helpful tool to visualize the coverage of specific variants.

### **Whole Genome Sequencing (WGS)**

Whole Genome Sequencing was used in **Project 3**, as no potential candidate gene was identified by WES. WGS was performed as previously described<sup>186</sup> on three affected siblings (II.2, II.4, II.7) and two healthy brothers (II.1, II.6) by the *Centre National de Recherche sur le Génome Humain* (CNRGH, Evry France). Genomic DNA was used to prepare a library for WGS using the Illumina TruSeq DNA PCR-Free Library Preparation Kit. Then a normalization and quality control were performed. In order to reach a sequencing depth of 30x, qualified libraries were sequenced on a HiSeq2000 platform (Illumina Inc., CA, USA), as paired-end 100 bp reads. Throughout the sequencing run, quality parameters were assessed. To generate FASTQ files for each sample standard bioinformatics analysis of sequencing data were performed according to the Illumina pipeline. The sequence reads were aligned to the reference sequence of the human genome (GRCh37) using the Burrows-Wheeler Aligner (BWA V7.12)<sup>187</sup>. The UnifiedGenotyper and HaplotypeCaller modules of the Genome Analysis ToolKit (GATK)<sup>188</sup>, Platypus (<https://github.com/andyrimmer/Platypus>) and Samtools<sup>189</sup> were used for calling both SNV and indel. Structural Variations (SV) were called using SoftSV<sup>190</sup>. Moreover, each known cataract and deafness genes were visually inspected with IGV<sup>191</sup>.

Annotation and ranking of SNVs/indels and structural variations were performed respectively by VaRank<sup>192</sup> (in combination with Alamut Batch, Interactive Biosoftware, Rouen, France) and by AnnotSV<sup>193</sup>. Variant effect on the nearest splice site was predicted using MaxEntScan<sup>194</sup>, NNSplice<sup>195</sup> and Splice Site Finder<sup>196</sup>. The following filtering criteria were applied to filter out non-pathogenic variants: (1) variants represented with an allele frequency of more than 1 % in public variation databases were excluded (Used databases: our internal exome database, 1000 Genomes<sup>197</sup>, the gnomAD database<sup>198</sup> and the DGV<sup>199</sup>), (2) variants in 5' and 3' UTR, downstream, upstream, intronic and synonymous locations without pathogenic prediction of local splice effect. The *PSMC3* nomenclature is based on the accession number NM\_002804.4 from the RefSeq database<sup>200</sup>. Genomic coordinates are defined according to GRCh37/hg19 assembly downloaded from the University of California Santa Cruz (UCSC) genome browser.<sup>201</sup>

From 4,990,312 to 5,165,496 genetic variants (SNV/indel/SV) were identified per individual from the WGS analysis (supplemental Table S3.2). Bioinformatic analyses highlighted 6 homozygous variations in 5 genes (see supplemental Method S3.2).

### 3.2.3 Cell culture techniques

#### **Cultivation of cells**

Primary skin fibroblasts of the patient and healthy control individuals of the same age and gender were obtained from skin biopsies as previously described.<sup>171</sup> Fibroblasts were cultivated at 37 °C and 5 % CO<sub>2</sub> in DMEM + GlutaMAX (Thermo Fisher Scientific) supplemented with 10 % FCS and 1 % anti-anti.

#### **Ciliogenesis**

For ciliogenesis, 1 x 10<sup>5</sup> fibroblast cells were plated in each well of an 8-well Nunc Lab-Tek Chamber Slide 177445 (Thermo Fisher Scientific). Once confluent, the cells underwent a 48-hour serum deprivation before fixation. The deprivation was executed by cultivating human fibroblasts in DMEM + GlutaMAX supplemented with 1 % anti-anti, without FCS.

### 3.2.4 Molecular biology techniques

#### **Amplification and analysis of DNA fragments**

The amplification of specific nucleic acid sequences was performed using the polymerase chain reaction (PCR). The PCR reaction was set-up as follows: 50 ng template DNA, 0.4 – 1 µM forward and reverse primer (Table 1), 1 U/µl Taq Polymerase (Sigma-Aldrich), 1X PCR buffer (Sigma-Aldrich), 1.5 mM MgCl<sub>2</sub> and 1mM dNTPs. For GC-rich sequences, 1 M Betaine and 3 % DMSO were added. PCR was performed in the Mastercycler ep gradient S (Eppendorf) with the following cycling conditions: an initial denaturation at 94 °C for 5 min, 30-40 denaturation-annealing- elongation cycles at 94 °C for 30 s, primer-specific annealing temperature for 30 s and 72 °C for 30 s, followed by a final elongation step at 72 °C for 10 min. Then, the PCR-products were separated according to their size by agarose gel electrophoresis on a 1.5 % agarose gel at 200 V in Tris/Borate/EDTA (TBE) running buffer (Euromedex). The progress of the electrophoresis was monitored using colored dye. For the visualization of the bands ethidium bromide was used and the GeneRuler DNA Ladder Mix (Thermo Fisher Scientific) was used to determine the size of the bands.

#### **Sample preparation of Sanger Sequencing**

Sanger sequencing was performed to identify/confirm variants in nucleic acid sequences of specific genes. The PCR products amplified as described above were purified to remove PCR components, like primers and unincorporated dNTPs. For this purpose, PCR reactions were loaded on a PCR Cleanup

Filter Plate (Millipore) and filtered with a vacuum manifold (Millipore) until the wells were dry. Then, water was added to the wells and agitated with the DNA for 20 min by shaking. Finally, the dissolved DNA and an aliquot of 10  $\mu$ M forward or reverse primer were sent to GATC Biotech (Germany) for sequencing and capillary electrophoresis. The software SeqScape v2.5 (Thermo Fisher Scientific) was used to align the sequences to the reference sequence GRCh37/hg19 and to visualize the chromatographs.

### **Extraction of total RNA and cDNA Synthesis**

Total RNA from patient fibroblasts was extracted with the RNeasy Kit (Qiagen). RNA integrity was evaluated by agarose gel electrophoresis and RNA concentration was determined with the Qubit RNA assay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. For the reverse transcription of the RNA into cDNA, 1  $\mu$ g RNA was used. The cDNA synthesis was carried out using the iScript cDNA Synthesis Kit (BioRad), which uses a blend of oligo (dT) and random hexamer primers, according to the manufacturer's instructions.

### **Real-time quantitative RT-PCR (qRT-PCR)**

In contrast to conventional end-point PCR, qPCR allows the detection of the accumulation of amplified products in real-time and can be used to quantify the expression of specific genes. Primers were designed (Table 3) and the best qPCR conditions were determined. To optimize the qPCR assay a range of annealing temperatures around the calculated  $T_m$  of the primers were tested. A standard curve using serial dilutions of template cDNA was performed to verify the efficiency of the primers. The specificity of the primers was evaluated using a melt curve. The reactions were prepared using the iQ SYBR Green Supermix (BioRad) according to the manufacturer's instructions using 2  $\mu$ l of cDNA. Reactions were set up in triplicate. The cycling parameters were as follows: 95 °C for 5 min, 45 cycles of 95 °C for 15 s and 60 °C for 35 s, followed by the generation of melt curves by heating in 0.5 °C increments (5 s/step) for the temperature range 65 to 95 °C. Gene expression levels were quantified relative to the reference genes *GAPDH* and *ACTB* ( $\beta$ -Actin) using the efficiency-corrected comparative cycle threshold ( $C_T$ ) method using the CFX Manager Software V.1.5 (BioRad).

### **Transcriptomic studies/ RNA Sequencing**

Transcriptomic studies were performed as previously described.<sup>202,203</sup> RNA samples were extracted from fibroblast of individual A.II-1, C.II-1, C.II-2 and from 3 controls in 2 conditions (ciliary and non-ciliary condition see cell culture section) using the RNeasy Mini Kit Qiagen®. A DNase I recombinant treatment (Sigma-Aldrich®) was performed. The integrity of the RNA was visualized on a 1 % agarose gel.<sup>204</sup> The quality and quantity of the RNA was assessed using a Nanodrop®. By running samples on a

RNA 6000 Nano Chip on the Bioanalyzer (Agilent Technologies), the integrity and quality of the RNA were evaluated again. The RNA integrity number (RIN) should be equal or above 8. Library preparation was performed at the Institute of Genetics and Molecular and Cellular Biology (Strasbourg, France) at the GenomEast platform, using the TruSeq® RNA sample preparation v2 protocol (Illumina) starting from 1 µg of extracted total RNA. 2x100bp paired-end sequenced was performed, with 2 samples per lane on an Illumina HiSeq4000 sequencer.”<sup>203</sup>

Reads were preprocessed in order to remove adapter and low-quality sequences (Phred quality score below 20). After this preprocessing, reads shorter than 40 bases were discarded for further analysis. These preprocessing steps were performed using cutadapt<sup>205</sup> version 1.10. Reads were mapped to rRNA sequences using bowtie<sup>206</sup> version 2.2.8, and reads mapping to rRNA sequences were removed for further analysis. Reads were mapped to spike sequences using bowtie<sup>206</sup> version 2.2.8, and reads mapping to spike sequences were removed for further analysis. Reads were mapped onto the hg19 assembly of Homo sapiens genome using STAR<sup>207</sup> version 2.5.3a. Coverage was computed for each gene percentile using geneBodyCoverage from RSeqQC<sup>208</sup> version 2.6.4. The median of transcript integrity numbers (TIN)<sup>209</sup> computed across all transcripts with at least 10 mapped reads. TIN is a metric dedicated to capture the uniformity of coverage for a given transcript, and thus is a measure of RNA integrity. TIN score is the percentage of transcript that has uniform read coverage and ranges from 0 (the worst) to 100 (the best). Gene expression quantification was performed from uniquely aligned reads using htseq-count<sup>210</sup> version 0.6.1p1, with annotations from Ensembl version 75 and \union" mode2. Only non-ambiguously assigned reads have been retained for further analyses. Read counts have been normalized across samples with the median-of-ratios method proposed by Anders and Huber<sup>211</sup>, to make these counts comparable between samples. The generated data were analyzed and visualized to + check the quality of the data and to eventually remove data with a bad quality. Especially, PCA was built after stabilizing variance using the regularized log transformation method or RSeqQC was used to do some quality controls such as gene body coverage. Only genes with | log2 fold-change | > 1 were considered, as we expected to identify a large amount of change between transcriptomic results from one patient compared to other individuals (indirect consequences of the pathogenic mutation or unrelated events due to variability between humans). To avoid as much as possible false positive variants, we filtered out all genes with a Benjamini and Hochberg adjusted p-value above 0.05.

Alternative splicing analysis was performed by JunctionSeq<sup>212</sup> version 1.6.0 and LeafCutter<sup>213</sup> version 0.2.7. For JunctionSeq, exon and splice junction counts, as well as novel splice junction counts, were generated using QoRTs<sup>214</sup> version 1.2.42. Differential usage of exons and splice junctions was tested with runJunctionSeqAnalyses function by using the default parameters. Concerning LeafCutter, novel and existing alternative events were detected by using default parameters except the following during



the intron clustering step in order to adjust for rare diseases:  $\text{mincluratio}=0.00001$  and  $\text{maxintronlen}=500000$ .

### **Native-PAGE and proteasome in-Gel peptidase activity assay**

Patient and control fibroblasts were lysed in equal amounts of TSDG buffer (10 mM Tris pH 7.0, 10 mM NaCl, 25 mM KCl, 1.1 mM  $\text{MgCl}_2$ , 0.1 mM EDTA, 2 mM DTT, 2 mM ATP, 1 mM  $\text{NaN}_3$ , 20 % Glycerol) prior to protein quantification using a standard BCA assay (Pierce) following the manufacturer's instructions. Twenty micrograms of whole-cell extracts were loaded on 3-12 % gradient native-PAGE gels (Invitrogen) and subsequently subjected to a 16-h electrophoresis at 45 V using a running buffer consisting of 50 mM BisTris/Tricine, pH 6.8. Chymotrypsin-like activity of the separated proteasome complexes was then measured by incubating 0.1mM of the suc-LLVY-AMC substrate (Bachem) at 37 °C for 20 min in a final volume of 10 ml of overlay buffer (20 mM Tris, 5 mM  $\text{MgCl}_2$ , pH 7.0. Proteasome bands were then visualized by exposure of the gel to UV light at 360 nm and detected at 460 nm using an Imager

## 3.2.5 Biochemistry techniques

### **Measurement of proteasome activity in crude extracts**

Ten micrograms of whole-cell extracts deriving from control and patient fibroblasts were tested for chymotrypsin-like activity by exposing the cell lysates to 0.1 mM of the Suc-LLVY-AMC peptide. Assays were carried out in a 100  $\mu\text{l}$  reaction volume of ATP/DTT lysis buffer at 37 °C. The rate of cleavage of fluorogenic peptide substrate was determined by monitoring the fluorescence of released aminomethylcoumarin using a plate reader at an excitation wavelength of 360 nm and emission wavelength of 460 nm over a period of 120 min.

### **Protein extraction**

**Project 1** | Total protein from patient fibroblasts was extracted with non-denaturing lysis buffer. 200-400  $\mu\text{l}$  of lysis buffer (20 mM Tris HCl (pH 8), 137 mM NaCl, 10 % Glycerol, 1 % Nonidet P-40 (NP-40), 2mM EDTA and 1x Cocktail Roche (Protease Inhibitor Cocktail)) was added to a cell pellet containing approximately  $4 \times 10^6$  cells. After pipetting the suspension ten times up and down, the suspension was agitated for 20 min at 4 °C and centrifuged for 30 min at full speed at 4 °C. The supernatant was transferred into a new Eppendorf tube and proteins were quantified with the Qubit protein assay kit (Thermo Fisher Scientific) according to the manufacturer's instructions.

**Project 2 and 3** | Skin fibroblasts from controls and the affected individual were recovered in ice cold RIPA buffer (50 mM Tris pH 7.5, 150 mM NaCl, 2 mM EDTA, 1 mM N-ethylmaleimide, 10  $\mu\text{M}$  MG-132,

1 % NP40, 0.1 % SDS) complemented with protease inhibitors ('Complete EDTA-free'; Roche Diagnostics, 1 tablet in 10ml buffer) and 25mM N-ethylmaleimide (NEM, diluted freshly in ethanol to prevent artifactual deubiquitination), left on ice for 45min, centrifuged for 10min at 12000rpm, supernatant recovered, proteins were quantified with the Qubit protein assay kit (Thermo Fisher Scientific) according to the manufacturer's instructions and 5X Laemmli buffer added.

### **Western Blot**

Western Blot is a method, which allows the transfer of proteins from a gel matrix onto a membrane via an electric field and the subsequent immunological detection of proteins by antibodies. In this work, 50 µg of total protein was supplemented with 1x Laemmli Sample Buffer (BioRad) and denatured for 15 min at 95 °C and at 1400 rpm. Then, proteins were separated according to their molecular weight at 210 V using a 10 % Mini-PROTEAN TGX gel (BioRad) and 10X TGS (BioRad) running buffer. The separated proteins were transferred onto a polyvinylidene difluoride (PVDF) membrane using the Trans-Blot Turbo Mini PVDF or cellulose Transfer Pack and Trans-Blot Turbo Blotting System (BioRad). The transfer was verified by dyeing the PVDF/ cellulose membrane with Ponceau stain (Sigma Life Science) after the transfer step. To prevent nonspecific binding of the antibodies, the membrane was then incubated for 4 h in 5 % milk in Tris-Buffered Saline and Tween 20 (Sigma-Aldrich) (TBST). For the specific detection of the protein of interest, the membrane was incubated overnight, under agitation at 4 °C with the primary antibody (Table 9). The membrane was then washed six times for 5 min in TBST, incubated for 1 h with an appropriate secondary antibody conjugated to a horseradish peroxidase (HRP) (Table 9) and washed another two times with TBS. Finally, proteins were detected with the SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) or the Novex Chemiluminescent Substrate (Thermo Fisher Scientific) according to the instructions in the kit's manuals using the Molecular Imager Gel Doc XR System (BioRad). The molecular weight of the bands was estimated using the Precision Plus Protein WesternC Standard (BioRad). Bands were quantified relative to the total amount of protein loaded (stainfree) or relative to a housekeeping gene using Image Lab. The mean of independent experiments was calculated and represented as a histogram and a student test performed to determine the significance.

### **Co-immunoprecipitation and mass-spectrometry analysis.**

**Project 3 |** PSMC3 protein and its partners were immunoprecipitated from patient's fibroblasts using magnetic micro particles (MACS purification system, Miltenyi Biotec) according to the manufacturer's instructions and as previously described.<sup>215</sup> Briefly, PSMC3 complexes were captured with an anti-PSMC3 antibody (Abcam ab171969) and the target and its associated proteins were purified on the protein G microbeads (Miltenyi Biotec). Proteins were eluted out of the magnetic stand with 1x laemmli buffer (Biorad). To optimize reproducibility, co-immunoprecipitation experiments were

carried out in affinity triplicates on exactly 1.1 mg of proteins for each sample. For negative controls, halves of each sample were pooled and immunoprecipitated with the protein G beads, omitting the antibody.

Protein extracts were prepared as described in a previous study.<sup>216</sup> Samples were precipitated with 0.1 M ammonium acetate in 100 % methanol. Proteins were resuspended in 50 mM ammonium bicarbonate. This was followed by a reduction-alkylation step (dithiothreitol 5 mM – iodoacetamide 10 mM) and the overnight digestion of proteins with sequencing-grade porcine trypsin (1:25, w/w, Promega, Fitchburg, MA, USA). The resulting vacuum-dried peptides were resuspended in water containing 0.1 % (v/v) formic acid (solvent A). One fifth of the peptide mixtures were analyzed by nanoLC-MS/MS on an Easy-nanoLC-1000 system coupled to a Q-Exactive Plus mass spectrometer (Thermo-Fisher Scientific, USA) operating in positive mode. Five microliters of each sample were loaded on a C-18 precolumn (75  $\mu$ m ID  $\times$  20 mm nanoViper, 3  $\mu$ m Acclaim PepMap; Thermo) coupled with the analytical C18 analytical column (75  $\mu$ m ID  $\times$  25 cm nanoViper, 3  $\mu$ m Acclaim PepMap; Thermo). Peptides were eluted with a 160 min gradient of 0.1 % formic acid in acetonitrile at 300 nL/min. The Q-Exactive Plus was operated in data-dependent acquisition mode (DDA) with Xcalibur software (Thermo-Fisher Scientific). Survey MS scans were acquired at a resolution of 70K at 200 m/z (mass range 350-1250), with a maximum injection time of 20 ms and an automatic gain control (AGC) set to 3e6. Up to 10 of the most intense multiply charged ions ( $\geq 2$ ) were selected for fragmentation with a maximum injection time of 100 ms, an AGC set at 1e5 and a resolution of 17.5K. During the peak selection process, a dynamic exclusion time of 20 s was applied.

MS data were searched against the Swissprot database (release 2019\_05) with Human taxonomy. We used the Mascot algorithm (version 2.5, Matrix Science) to perform the database search with a decoy strategy and search parameters as follows: carbamidomethylation of cysteine, N-terminal protein acetylation and oxidation of methionine were set as variable modifications; tryptic specificity with up to three missed cleavages was used. The mass tolerances in MS and MS/MS were set to 10 ppm and 0.05 Da respectively, and the instrument configuration was specified as “ESI-Trap”. The resulting .dat Mascot files were then imported into Proline v1.4 package (<http://proline.profi-proteomics.fr/>) for post-processing. Proteins were validated with Mascot pretty rank equal to 1.1 % FDR on both peptide spectrum matches (PSM) and protein sets (based on score). For the quantification of each protein in the different samples, the total number of MS/MS fragmentation spectra was used.

Data collected from multiple experiments were used for the statistical analysis of the co-immunoprecipitation data. We compared them against the negative control IPs using a homebrewed R package as described previously.<sup>217</sup> except that that the DESeq normalisation method was used to calculate the size factor used to scale the samples (i.e. median of ratios method<sup>211</sup>). The package

calculates the fold change and an adjusted P-value corrected by Benjamini–Hochberg for each identified protein (and visualizes the data in volcano plots).

### **Immunohistochemistry**

Immunohistochemistry allows the localization of proteins within cells.  $8 \times 10^4$  cells were grown in each well of Nunc Lab-Tek Chamber Slides (Thermo Fisher Scientific) to the desired confluence before fixation. The cells were rinsed with PBS and fixed with 4 % paraformaldehyde for 30 min at room temperature. After three washing steps with PBS, cells were permeabilized for 10 min with 0.2-0.5 % Triton X-100, washed three times with PBS and blocked for 1 h in 20 % FBS. The overnight incubation with the primary antibody (Table 2) was performed at 4 °C. Cells were then washed six times in PBS, incubated for 1 h with a secondary antibody conjugated to a fluorophore (Table 3), washed six times in PBS and stained with 1:2000 Hoechst33258 Pentahydrate (Life Technologies) for 10 min. The slides were mounted with Vectashield (Vector Laboratories).

### **Imaging**

**Project 1 |** Images of the cells were taken with the fluorescence microscope (Zeiss Axio Observer D1). The length of primary cilia was measured on the pictures using Image J.

**Project 2 |** A fluorescence (Zeiss Axio Observer D1) or confocal (Zeiss LSM700, Plateforme Microscopie et Imagerie (IBMP, Strasbourg, France)) microscope, X40 objective was used to observe cells. Length of actin fibers was measured on the pictures using ImageJ.

**Project 3 |** A fluorescence (Zeiss Axio Observer D1) or confocal (Zeiss LSM700, Plateforme Microscopie et Imagerie (IBMP, Strasbourg, France)) microscope, X40 objective was used to observe cells.

### 3.3 Methods carried out in the Institute of Toxicology and Genetics, Karlsruher Institut für Technologie (KIT)

#### 3.3.1 Zebrafish (*Danio rerio*) handling

##### **Maintenance and husbandry**

In this study the zebrafish wild-type line AB202 (University of Oregon, Eugene), mutant lines (Table 7) and the transgenic line gSAIzGFF539A (SOKENDAI, the Graduate University for Advanced Studies, Mishima, Japan) were used and maintained at 28 °C under a 14 hour light and 10 hour dark cycle as described previously.<sup>218,219</sup> Fish were fed with commercial food and Artemia. When fish reached sexual maturity, zebrafish couples were transferred to breeding tanks the day before and crossed after the beginning of the next light cycle. Fertilized zebrafish eggs were raised at 28.5 °C in 1× Instant Ocean salt solution (Aquarium Systems, Inc.). Zebrafish husbandry and experimental procedures were performed in accordance with German animal protection regulations (Regierungspräsidium, Karlsruhe, Germany, AZ35-9185.81/G-137/10).

##### **Microinjections**

Zebrafish embryos were injected in the morning between 9 and 11 am. One day before sexual mature zebrafish were transferred into breeding tanks. A transparent disc or net separated female and male fish from each other. In the morning the separation between both genders was removed. The fertilization of the eggs followed only a few minutes later. The eggs were collected using a fine net and transferred to petri dishes. The injection needles were prepared using the Flaming/Brown needle puller. The needles were filled with 3-6 µl. Using a gas-driven microinjector system (Tritech) zebrafish embryos were injected in the single-cell or two-cell stage either into the cell or into the yolk directly below the cell. To ensure the same amount of injection solution/embryo 0.1 % of phenol red was added to each sample. After injection, the embryos were kept at 28.5 °C and unfertilized or dead embryos were sorted out.

For CRISPR/Cas9 injections 300 ng/µl of Cas9 protein (GeneArt Platinum Cas9 Nuclease, Invitrogen) and 300 ng/µl guide RNA were combined and injected into one-two cell stage eggs.

The morpholino concentration that needs to be injected in one cell stage embryos can differ enormously from morpholino to morpholino. Prior to be able to observe a gene specific phenotype different concentration were tested. The most suitable concentration for *prmt9-mo* was 0.3 mM (**project 1**), 0.03 mM for *myh10-mo*, 0.06 mM for *myh10-mo2* (**project 2**) and 0.2 mM for *psmc3-mo* (**project 3**). The corresponding control morpholinos were injected with the same concentrations.

For rescue experiments, mRNA of zebrafish *psmc3* was synthesized using the mMMESSAGE mMACHINE system (Ambion). Zebrafish *psmc3* mRNA was injected at a final concentration of 10 ng/µl.

### 3.3.2 Generation of CRISPR/Cas9 knockout mutants

#### Guide RNA design

To knockdown a specific gene in zebrafish the CRISPR/Cas9 system was used. To disrupt the function of a gene, a guide RNA targeting one of the first exons was chosen to introduce a frameshift and thus an early premature stop codon. The 20 bp binding sequences for CRISPR guide RNAs were chosen from the ChopChop software (Table 4).

#### Guide RNA synthesis

CRISPR guide RNAs were synthesized using a cloning-free guide RNA synthesis method.<sup>220</sup> An oligonucleotide sequence containing a T7 promotor (5'-TAATACGACTCACTATA-3'), the gene-specific sequence chosen using Chopchop (without the PAM sequence) and a complementary sequence (5'-GTTTTAGAGCTAGAA-3') was ordered from Metabion and annealed to a constant oligonucleotide (5'-AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTATTCTAGCTCTAAAAC-3'). To anneal both oligos, they were incubated for 5 min at 95 °C in a block heater (Stuart Equipment, Staffordshire, UK). The block heater was switched off and the samples were kept in the instrument until the machine was cooled down.

##### Annealing of oligos:

Gene-specific oligo (100 µM)	1 µl
Constant oligonucleotide (100 µM)	1 µl
<u>Water</u>	<u>8 µl</u>
Total:	10 µl

To fill in the single stranded overhangs the following components were added to the annealed oligonucleotide sequences and incubated for 20 min at 12 °C.

##### To the annealed oligos above, add T4 DNA polymerase to fill-in:

dNTPs (25 mM)	1 µl
10x NEB buffer 2	2 µl
100x NEB BSA	0,2 µl
T4 NEB DNA polymerase	0,5 µl
<u>Water</u>	<u>6,3 µl</u>
Total:	10 µl

The guide RNA template was purified with the Nucleospin Gel and PCR Clean-up kit (Macherey-Nagel) and the guide RNA was transcribed with the MEGAscript T7 Transcription Kit (Ambion). Both kits were used according to the manufacturer's instructions. The RNA was aliquoted and stored at -80 °C.

#### Guide RNA injection (F0-Generation)

For CRISPR/Cas9 injections an equal amount of Cas9 protein (GeneArt Platinum Cas9 Nuclease, Invitrogen) and guide RNA (300-400 ng/µl) were combined and injected into one to two cell stage eggs

as described before. After 24 hpf, the chorion of 10-15 embryos was manually removed with forceps, embryos were anesthetized with 0.02 % tricaine and the DNA of all 10-15 embryos was extracted in a single tube (see chapter **DNA extraction**). A PCR was performed using specific primers encompassing the binding sequence of the guide RNA (table 4) and a Sanger Sequencing of the PCR product was conducted by Microsynth (see chapter **PCR, gel electrophoresis**). Injected embryos (F0-Generation) are often genetically mosaics. In case a guide RNA efficiently cuts its target, the Sanger sequencing electropherogram shows a clear frameshift just in front of the PAM sequence. Embryos injected with efficient guide RNAs were raised until sexual maturity (founders).

### **Generation of crispants**

In cases of highly efficient gRNAs, the mosaic F0-Generation (also called crispants) has been shown to recapitulate mutant phenotypes successfully.<sup>221-223</sup> In **project 3**, this strategy was used to recapitulate the eye and ear phenotype seen in affected patients. The web interface PCR-F-Seq q (<http://iai-gc-server.iai.kit.edu>)<sup>224</sup> was used to quantify the cutting efficiency of guide RNAs (supplemental Figure S3.13).

### **Fin biopsy/ Genotyping**

To identify potential founders for the next generation, adult zebrafish (e.g. F0-generation) were genotyped. After anesthetizing them with tricaine, fish were placed in a lateral position and a small piece of the fin was cut off with a scalpel. The DNA was extracted as described in the following section (see **Genomic DNA extraction**). A PCR was performed using specific primers encompassing the binding sequence of the guide RNA (table 4) and a Sanger Sequencing of the PCR product was conducted by Microsynth (see chapter **PCR, gel electrophoresis**).

### **Generation of stable knockout lines**

Genetically mosaic founders were crossed with wildtype fish. The heterozygous F1-Generation was genotyped. Only fish carrying a frameshift mutation leading to a premature stop codon were chosen to generate stable knockout lines. The F1-Generation was crossed out to obtain more heterozygous carrier of the same mutation and to get rid of potential nonspecific mutations. To generate homozygous mutants, an incross of heterozygous mutants was carried out.

### 3.3.3 Molecular biology methods

#### Genomic DNA extraction

DNA was extracted by transferring the tissue (whole embryo or fin biopsy) into a PCR tube with 100  $\mu$ l 50 mM NaOH, followed by an incubation of 20 min at 95 °C and 5 min at 4 °C (Hotshot). Afterwards, 10  $\mu$ l of 1M Tris HCl (pH 7.5) was added to neutralize the solution. In order to identify genetically modified fish, 5  $\mu$ l of the solution was used to carry out a PCR (25  $\mu$ l PCR reaction).

#### PCR and agarose gel electrophoresis

PCRs were performed to amplify specific gene sequences, e.g. for genotyping. A PCR reaction was set-up as follows:

gDNA/ cDNA	10 $\mu$ l
5x GreenGpTaq Reaction buffer	10 $\mu$ l
GoTaq G2 DNA Polymerase	0,4 $\mu$ l
25 mM DNTPs	0,4 $\mu$ l
10 $\mu$ M Primer (forward + reverse)	0,4 $\mu$ l
25 mM MgSO <sub>4</sub>	3 $\mu$ l
DMSO	2,5 $\mu$ l
H <sub>2</sub> O	23,3 $\mu$ l
Total:	50 $\mu$ l

PCR was performed in a thermocycler with the following cycling conditions: an initial denaturation at 94 °C for 5 min, 40 denaturation-annealing- elongation cycles at 94 °C for 30 s, primer-specific annealing temperature for 30 s and 72 °C for 30-90s, followed by a final elongation step at 72 °C for 10 min. The elongation time was adapted to the length of the PCR product (1kb/minute).

Then, the PCR-products were separated according to their size by agarose gel electrophoresis on 1.5-2 % agarose gel at 120 V in TAE running buffer. The progress of the electrophoresis was monitored using colored dye. For the visualization of the bands, ethidium bromide (3  $\mu$ l/100 ml) was added to the agarose gels. The GeneRuler, DNA Ladder Mix/ 50bp (Thermo Fisher Scientific) was used to determine the size of the bands. Images of the gel were taken using the Gel documentation system (GenoPlex, VWR).

#### Isolation of DNA from agarose

To isolate DNA from agarose, under UV-light, the appropriate band was cut out with a scalpel. For the purification of the DNA the Nucleospin Gel and PCR Clean-up kit (Macherey-Nagel) was used following the instructions of the supplier.



### **Cloning**

Cloning was performed to generate DNA templates for the transcription of RNA DIG probes, full-length mRNAs for rescue experiments or short mRNA sequences bound to GFP to check the efficiency of morpholino binding.

#### Restriction and Ligation of plasmid DNA

Plasmid DNA was digested with restriction enzymes and the corresponding buffer according to the supplier's instructions. Approximately one unit of enzyme was used to digest 1 µg of plasmid DNA. Digestion was performed for 1h or overnight. For the ligation, a vector to insert ratio of 1:3 was chosen. The amount of insert was calculated with the help of the online ligation calculator of the university of Düsseldorf ([http://www.insilico.uni-duesseldorf.de/Lig\\_Input.html](http://www.insilico.uni-duesseldorf.de/Lig_Input.html)).

#### Bacterial Transformation.

To increase the quantity of (freshly ligated) plasmid DNA, plasmids were transferred into bacteria (XL1 Blue competent cells). 50 µl of bacteria-solution was thawed on ice, combined with 20 µl of the ligation product and incubated for 20-30 min on ice. A heat shock was performed for 40 seconds at 42 °C, followed by an incubation for 5 min on ice. The bacterial solution was plated on LB agar plates containing a suitable antibiotic and incubated overnight at 37 °C. The next day, single bacterial colonies were picked with pipette tips, transferred into 50 ml LB medium in Erlenmeyer flasks and incubated at 37 °C overnight in a New Brunswick Innova 44 Incubator Shaker.

#### Isolation of plasmid DNA

Plasmid DNA was isolated from overnight-cultures following the suppliers instructions using the Mini and Midi Plasmid preparation Kit of Macherey-Nagel. The concentration of plasmid DNA was measured using a Nanodrop.

### **Extraction of total RNA**

For total RNA extraction from whole zebrafish embryos, 1 ml of Trizol reagent (Invitrogen) was added to 50 embryos. The tissue was mechanically disrupted using a syringe and incubated for 5 min at room temperature. 200 µl chloroform was added to the homogenized solution, vortexed for 15 seconds, incubated for 2 min at room temperature and centrifuged for 20 min at 13,000 rpm at 4 °C. The upper aqueous phase was transferred into a clean Eppendorf tube, 500 µl of isopropanol was added, mixed by vortexing and incubated for 1h at -80 °C. Afterwards, the solution was centrifuged for 30 min at 4 °C. The supernatant was removed and 500 µl of 75 % EtOH was added to wash the pellet. After a centrifugation step for 5 min, at 13,000 rpm at 4 °C the pellet was air dried for 5-10 min at room temperature and then resuspended in 100 µl RNase-free H<sub>2</sub>O. The RNA integrity was evaluated by

agarose gel electrophoresis and RNA concentration was determined with the Nanodrop ND-1000 (Peqlab).

### **cDNA synthesis for RT PCR**

Prior to cDNA synthesis, RNA was digested with DNase to prevent genomic contamination. For this, to 1 µg of RNA the following products were added:

1 µg RNA	x µl
RQ1 RNase free DNase 10x buffer (Promega)	2 µl
RQ1 RNase free DNase (Promega)	2 µl
<u>Nuclease free H<sub>2</sub>O</u>	<u>x µl</u>
Total:	20 µl

The solution was incubated for 30 min at 37 °C. RQ1 DNase stop solution was added, followed by an incubation step of 10 min at 65 °C to terminate the reaction. For the cDNA synthesis, the Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Fisher Scientific) was used. The following components were combined in a PCR tube (on ice):

5x Reaction Mix	4 µl
Maxima Enzyme Mix	2 µl
2 µg template RNA	x µl
<u>Nuclease free H<sub>2</sub>O</u>	<u>x µl</u>
Total:	20 µl

The solution was mixed gently and then, incubated for 10 min at 25 °C, for 50 min at 50 µl and for 5 min at 85 °C in a thermocycler. Afterwards, 40 µl nuclease-free water was added and the cDNA was stored at -20 °C.

### **Quantitative qRT-PCR**

To check zebrafish mutants for NMD a quantitative RT-PCR was performed. Primers for the qRT-PCR were purchased from Metabion and are listed in Table 3. The reactions were prepared using the GoTaq 1-Step-RT-qPCR System (Promega) according to the manufacturer's instructions using 4 µl of cDNA per reaction.

2x Gotaq master mix	10 µl
10 µM Primer (forward + reverse)	2 µl
Rox (CXR reference dye)	0.2 µl
Nuclease free H <sub>2</sub> O	4 µl
<u>cDNA</u>	<u>4 µl</u>
Total:	20 µl

Reactions were set up in triplicate. The qPCR was performed with the StepOnePlus™ Real-Time PCR System (Thermo Fisher Scientific) using the following cycling parameters: 95 °C for 15 s, 60 °C for 30 s, 40 cycles of 95 °C for 15 s and 60 °C for 1 min, followed by the generation of melt curves by heating in

0.5 °C increments (5 s/step) for the temperature range 60 to 95 °C. Gene expression levels were quantified relative to the reference genes *β-actin* and *rpl13a* using the efficiency-corrected comparative cycle threshold ( $C_T$ ) method.

#### **mRNA transcription (rescue experiment)**

To exclude possible off-target effects caused by morpholino or CRISPR/Cas9 injection, in **project 3**, a rescue experiment was performed. Full-length *psmc3* was cloned into pCS2+ using EcoRI, linearized with NotI and *psmc3* full-length RNA synthesized using the mMESSAGING mMACHINE SP6 Kit (Ambion). Zebrafish *psmc3* mRNA was injected at a final concentration of 10 ng/μl.

### 3.3.4 Histological methods

#### **Immunohistochemistry**

Zebrafish embryos were raised in water supplemented with 0.003 % phenylthiourea (PTU) to suppress melanogenesis, dechorionated using forceps and fixed after 3 dpf in 4 % PFA for one hour at room temperature. After fixation, embryos were washed 4x 5min with PBS, permeabilized in Acetone for 7 min at -20 °C, washed 5 min in H<sub>2</sub>O, 3x 5min in PBS and blocked for two hours in 5 % BSA and 1% DMSO in 1xPBS. The overnight incubation with the primary antibody (Table 9) was performed at 4 °C. Embryos were then washed 6x 10 min in PTW (1x PBS, 0.1 % v/v Tween20), incubated for 2 hours with a secondary antibody conjugated to a fluorophore (Table 9) or with 1:500 Alexa Fluor 488 phalloidin (Life Technologies), washed 3x 10 min in PTW, stained with 1:1000 DAPI for 20 min and washed 3x 5 min in PBS. Immunostaining of whole zebrafish embryos after 5 dpf was performed as previously described.<sup>225</sup> Stainings were analyzed with the Leica TCS SP2 or SP5 confocal microscope.

#### ***In situ* hybridization (ISH)**

##### Synthesis and labelling of DIG RNA probes

Linearized plasmids served as templates for DIG RNA probe synthesis. The plasmids used for DIG RNA probe synthesis are listed in Table 6. For the digestion of plasmids, the following components were combined and incubated for 2 h at 37 °C:

5 μg DNA	x μl
Restriction enzyme	2 μl
10x buffer	5 μl
Nuclease free H <sub>2</sub> O	x μl
Total:	50 μl

Afterwards, nucleic acids were purified performing the phenol/chloroform extraction method. 50 μl were added to the solution containing the linearized plasmid. An equal volume of phenol/chloroform

(1:1) was added to the sample and vortexed. After a centrifugation for 15 min at 14,000 rpm at room temperature the supernatant (without the interphase) was transferred into a clean 1.5 ml Eppendorf tube. 100  $\mu$ l chloroform was added, the sample was vortexed and centrifuged for 15 min at 14,000 rpm at RT. Again, the supernatant (without the interphase) was transferred into a clean 1.5 ml Eppendorf tube and mixed with 100  $\mu$ l isopropanol by vortexing. After incubating the sample for 30 min at -80 °C, they were centrifuged for 15 min at 14,000 rpm at 4 °C. The supernatant was removed carefully and the pellet was washed by adding 75 % EtOH. After a centrifugation step for 5 min at 4 °C at full speed, the DNA was air dried for 5-10 min and resuspended in 16  $\mu$ l RNA-free water. 1  $\mu$ l of DNA was used to determine the DNA concentration.

For the RNA DIG probe synthesis, the following components were combined and incubated for 3 hours at 37 °C:

2 $\mu$ g linearized DNA	x $\mu$ l
RNasin	1.5 $\mu$ l
RNA polymerase Sp6	1.5 $\mu$ l
DIG labeling mix	2 $\mu$ l
5x transcription buffer	4 $\mu$ l
<u>Nuclease free H<sub>2</sub>O</u>	<u>x <math>\mu</math>l</u>
Total:	50 $\mu$ l

1  $\mu$ l DNase was added to the sample and incubated for 15 min at 37 °C to prevent genomic contamination. The synthesized RNA probe was diluted by adding 10  $\mu$ l DEPC water, 3.75  $\mu$ l NaOAc (pH 4.2) and 65  $\mu$ l EtOH. The sample was vortexed, incubated for 1 hour at -80 °C and centrifuged for 20 min at 14,000 rpm at 4 °C. The supernatant was removed carefully, the pellet was washed with 100  $\mu$ l 75 % EtOH and centrifuged for 5 min at 14,000 rpm at 4 °C. The pellet was air-dried for 5-10 min and resuspended in 11  $\mu$ l RNase-free water. The Nanodrop was used to assess the RNA concentration and the RNA quality was checked by running 1  $\mu$ l of the DIG labeled RNA on a 2 % agarose gel. To store the RNA DIG labeled RNA at -20 °C, 10  $\mu$ l of hybridization buffer (HYB) (Table 10) was added to the sample.

**Project 1 |** To analyze the expression pattern of *prmt9* in zebrafish embryos, a 552 bp *prmt9* fragment was amplified and cloned into the pGEMT-easy vector (Promega). The restriction enzyme Apa1 was used to linearize the plasmid and SP6 to transcribe the anti-sense RNA DIG probes.

**Project 3 |** To analyze the expression pattern of *psmc3* in zebrafish embryos, a 956 bp *psmc3* fragment was amplified and cloned into the pGEMT-easy vector (Promega). The restriction enzyme Apa1 was used to linearize the plasmid and SP6 to transcribe the anti-sense RNA DIG probes.

### Fixation and preparation of embryos for ISH

To suppress melanogenesis, zebrafish embryos were raised in water supplemented with 0.003 % phenylthiourea (PTU), dechorionated using forceps and fixed at 4 °C in freshly thawed BT-Fix. Overnight, embryos were transferred into MetOH. The next day embryos were rehydrated for 5 min in 75% MetOH in PTW, for 5 min in 50% MetOH in PTW, for 5 min in 25 % MetOH and four times for 5 min in PTW. Then, embryos were incubated in pre-cooled 80 % Acetone for 1 hour at -20 °C, washed for 5 min in PTW and digested with 10 µg Proteinase K. The Proteinase K treatment depended on the stage of the embryo (1 dpf for 5 min, 2 dpf for 45 min, 3 dpf for 2h). The supernatant was removed and the embryos were re-fixed for 20 min at 4 °C. Then, embryos were washed 3 times for 5 min in PTW and incubated for 3-4 hours at 70 °C in HYB buffer. DIG RNA labeled probes were diluted in HYB buffer (1:300) and incubated overnight at 70 °C. The next day, embryos were washed two times for 30 min in Wash Buffer 1, 15 min in Wash Buffer 2, 2 times for 30 min in Wash Buffer 3, 5 min in Wash Buffer 4 and three times for 5 min in PTW at room temperature. After washing the embryos once in Blocking Buffer, they were incubated for 1-2 hours at room temperature in fresh Blocking Buffer. The supernatant was removed and 300 µl of anti-Dig antibody diluted in Blocking Buffer (1:4000) was added to the embryos and incubated overnight at 4 °C. The next morning, embryos were washed five times for 5 min with PTW buffer at room temperature. Then, embryos were stained in darkness in Staining Buffer with 0.188 mg/ml NBT and 0.5 mg/ml BCIP in a 24 well plate. Once the staining was visible embryos were washed several times with PTW. To remove background a washing step with 100 % EtOH for 15 min was carried out, followed by several washing steps with PTW. Images of embryos were taken with the stereomicroscope MZ16 F (Leica) containing a DFC 320 (Leica) camera.

### **Imaging**

#### Imaging of whole embryos

Embryos were placed on a 1 % agarose plate and imaged using the stereomicroscope MZ16 F (Leica) containing a DFC 320 (Leica) camera.

#### Imaging of the zebrafish ear and eye

To examine the eye and ear phenotype in morphants and crispants, living embryos were anaesthetized with 0.0168 % tricaine and embedded them into 0.5 % low melting point agarose chilled to 37 °C in a lateral position with one eye/ear facing towards the objective. Confocal reflection microscopy was used to examine zebrafish eyes for abnormal light reflection evoked through cataract as previously described<sup>149</sup> and imaged with the Leica TCS SP2 confocal system with a 63x water immersion objective. Brightfield and fluorescence real time imaging of zebrafish ears were acquired using the Leica TCS SP5

confocal microscope with a 63x and 40x water immersion objective. The duration of the life imaging experiment was 16 hours. An image was taken every 20 min.

#### Birefringence

To check the muscle integrity of 3 dpf old zebrafish embryos, a birefringence assay was performed as previously described.<sup>226</sup> In a lateral position, embryos are visualized between two polarized filters and a dark background. Highly organized muscle fibers appear very bright in contrast to muscles with structural abnormalities. Images of embryos were taken with the DFC 320 (Leica) camera of the stereomicroscope MZ16 F (Leica).

### 3.3.5 Behavior test of adult zebrafish

#### **Social preference test of adult zebrafish**

For the social preference test we only used adult males. Standard 1L breeding tanks divided into two compartments with a clear barrier, were used, to separate a single fish (mutant or wildtype) from a conspecific group of five fish. After an acclimation period of 5 min, we recorded fish for 10 min. To analyze the swimming behavior, we used the tool ToxTrac.<sup>227,228</sup> Statistical analyzes were performed with the software Graphpad Prism 6.

**IV**

**Results and Discussion**

# Project 1

Based on:

Article in preparation, will be submitted to the American Journal of Human Genetics

## Biallelic variations in *PRMT9* delineate a novel syndromic form of intellectual disability

*Kröll-Hermi Ariane*<sup>1,2</sup>, *Stoetzel Corinne*<sup>1</sup>, *Schaefer Elise*<sup>1,2</sup>, *Scheidecker Sophie*<sup>1,3</sup>, *Prasad Megana*<sup>1</sup>, *Etard Christelle*<sup>4</sup>, *Kahrizi Kimia*<sup>5</sup>, *Geoffroy Véronique*<sup>1</sup>, *Schneider Aline*<sup>1</sup>, *Mattioli Francesca*<sup>6</sup>, *Plassard Damien*<sup>7</sup>, *Keime Céline*<sup>7</sup>, *Piton Amélie*<sup>4,6</sup>, *Wentzensen Ingrid*<sup>8</sup>, *Santiago-Sim Teresa*<sup>8</sup>, *Rustad Cecilie*<sup>9</sup>, *Oyvind B.*<sup>9</sup>, *Tveten Kristian*<sup>10</sup>, *Smeland Marie Falkenberg*<sup>10</sup>, *Alkhunaizi Ebba*<sup>11</sup>, *Li Chumei*<sup>12</sup>, *Kircher Susanne*<sup>13</sup>, *Laconne Franco*<sup>13</sup>, *Kolvenbach Caroline Maria*<sup>14</sup>, *Tasic Velibor*<sup>15</sup>, *Hildebrandt Friedhelm*<sup>14</sup>, *Keren Boris*<sup>16</sup>, *Whalen Sandra*<sup>16</sup>, *Roume Joelle*<sup>17</sup>, *van Haeringen Arie*<sup>18</sup>, *Najmabadi Hossein*<sup>5</sup>, *Strähle Uwe*<sup>2</sup>, *Dollfus Hélène*<sup>1,2,19</sup> and *Muller Jean*<sup>1,4</sup>

1. Laboratoire de Génétique médicale, UMR\_S INSERM U1112, IGMA, Faculté de Médecine FMTS, Université de Strasbourg, Strasbourg, France.
2. Service de Génétique Médicale, Institut de Génétique Médicale d'Alsace, Hôpitaux Universitaires de Strasbourg, Strasbourg, France.
3. Laboratoires de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France.
4. Karlsruhe Institute of Technology (KIT), Institute of Toxicology and Genetics (ITG), Eggenstein-Leopoldshafen, Germany.
5. Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.
6. Institut de Génétique et de Biologie Moléculaire et Cellulaire, INSERM U964, CNRS UMR 7104, Université de Strasbourg, 67400 Illkirch, France.
7. Plateforme GenomEast, IGBMC, Illkirch, France.
8. GeneDx Inc., Gaithersburg, MD 20877, USA.
9. Department of Medical Genetics, Oslo University Hospital, Oslo, Norway.
10. Department of Medical Genetics, University Hospital of North Norway, Tromsø, Norway.
11. Department of Clinical and Metabolic Genetics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada.
12. McMaster University Medical Center, Hamilton, ON, Canada.
13. Institute of Medical Genetics, Medical University of Vienna, Vienna, Austria.
14. University Children's Hospital, Medical Faculty of Skopje, Skopje, Macedonia.
15. Division of Nephrology, Department of Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, United States
16. AP-HP, Département de Génétique, Centre de Référence Maladies Rares "Anomalies du développement et syndromes malformatifs" Hôpital de la Pitié Salpêtrière, Paris, France.
17. Department of Clinical Genetics, Centre de Référence "AnDDI Rares", Poissy Hospital GHU PIFO, Poissy, France.
18. Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.
19. Centre de Référence pour les affections rares en génétique ophtalmologique, CARGO, Filière SENSGENE, Hôpitaux Universitaires de Strasbourg, 67091 Strasbourg, France.



## 4. Results

### 4.1 Project 1

Parts of the following sections (4.1.2-4.1.5) will be also published in the article: “**Biallelic variations in *PRMT9* delineate a novel syndromic form of intellectual disability**” that is currently in preparation and will be submitted to the “**American Journal of Human Genetics**”. The synopsis will give a brief summary about the project and points out my personal contribution to this work.

#### 4.1.1 Synopsis

Severe intellectual disability (ID) has a strong genetic component.<sup>229</sup> Although more than 700 ID genes have already been identified in Mendelian forms of isolated and syndromic ID, researchers estimate that this number will far exceed 1000 in the coming years. In project 1, I present a novel gene mutated in autosomal recessive syndromic intellectual disability.

To identify the pathogenic variation in an Algerian patient with severe ID, bilateral periventricular neuronal heterotopia, epilepsy, autism-spectrum behavior, and bilateral postaxial polydactyly, whole exome sequencing was performed. A single nucleotide deletion (c.545delT) in the coding sequence of the *PRMT9* gene was detected causing a frameshift and an early stop codon (p.Leu182Trpfs\*) that is predicted to give rise to a severely truncated protein. Using Sanger sequencing, it was shown that the mutation segregates with the disease in the family. *PRMT9* belongs to a protein family called the protein arginine methyltransferases (PRMTs). To date nine PRMTs are described and are categorized according to their methylation product. *PRMT1*, *PRMT2*, *PRMT3*, *PRMT4/CARM1*, *PRMT6* and *PRMT8* are Type I enzymes and able to monomethylate (MMA) or asymmetric dimethylate arginine residues (ADMA). *PRMT9* is as well as *PRMT5* a Type II enzyme and produces MMAs and symmetric dimethyl argenines (SDMAs). *PRMT7* is the only known Type III enzyme and creates only MMAs.<sup>167</sup> PRMTs have a wide spectrum of targets, such as histones, RNA-binding proteins, ion channels, transport proteins, scaffolding proteins, transcription factors, coactivators and proteins involved in neuronal transmission.<sup>230</sup> To date only one interaction partner of *PRMT9* is known, the splice factor *SAP145*, wherefore *PRMT9* is thought to play an important role in alternative splicing.<sup>3</sup> *PRMT9* was already proposed as a candidate gene for isolated ID.<sup>170</sup> However, only a single individual (no familial segregation available at that time) and the absence of functional experiments could not confirm this hypothesis.<sup>4</sup> Using GeneMatcher and Decipher, I was able to recruit eight additional families with patients carrying mutations in the gene *PRMT9*. All of them, presented a similar form of syndromic intellectual disability.<sup>5,6</sup> Beside ID, the most frequent symptoms were global developmental delay, epilepsy, impaired speech, autism and hypotonia.

As the index patient displayed bilateral postaxial polydactyly, renal anomalies, developmental delay and had a suspicion of RP, in its early infancy he was suspected to be affected by the ciliopathy BBS. Ciliopathies are pathologies associated with defects of the primary cilium, a cell organelle present on the surface of almost all cells of the body and playing a critical role in developmental and homeostatic signaling pathways.<sup>231</sup> In order to examine a potential effect of *PRMT9* deficiency on the ciliogenesis, skin fibroblasts of three patients and control cells were serum starved to induce primary cilia biogenesis. Indeed, primary cilia in all three patients were significantly longer than in control cells, suggesting a potential role of *PRMT9* during ciliogenesis.

Moreover, first experiments indicated that mutated forms of *PRMT9* might have an impact on alternative splicing. *PRMT9* is known to be responsible for the methylation of the splice factor *SAP145*, in Western Blot experiments, however, the methylated form of *SAP145* was only detectable in control cells, showing that the mutated forms of *PRMT9* of three different patients were not able to methylate its interaction partner.

To identify differences in alternative splicing and gene expression caused by *PRMT9* mutations, RNA sequencing was performed on the cells of three patients and three controls. To find a link between *PRMT9* and the patients' symptoms, I compared the list of genes assigned to databases on genes for intellectual disability, autism, and genes associated with primary cilia function and biogenesis. Interesting differences of gene expression were then validated via qPCR. For example, candidate genes for intellectual disability and autism, *PCDH10*, *SHROOM4* and *PLXDC2*, are expressed much higher in patients' cells than in control fibroblasts, which reinforces the hypothesis that *PRMT9* is a gene for intellectual disability. Induction of ciliogenesis by serum deprivation in patients' cells relative to control cells resulted in the deregulation of the mRNA expression of well-known cilia genes such as *GLI2* and *RPGRIP1L*, indicating that *PRMT9* could play a role during the biogenesis of the primary cilia and might have an impact on the hedgehog signaling pathway.

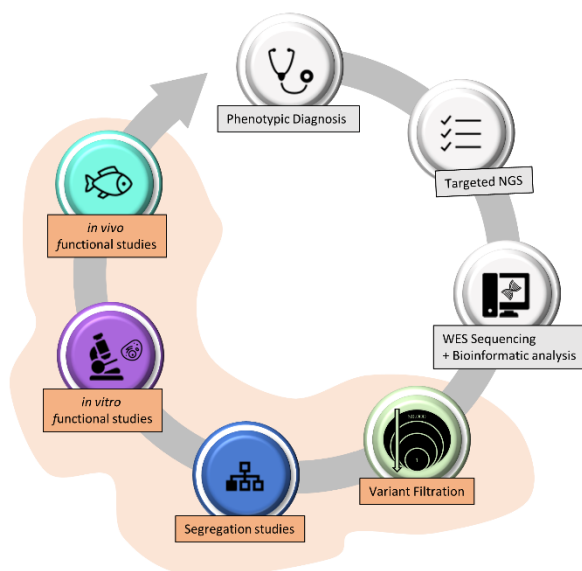
In addition, I used the CRISPR/Cas9 system to generate a zebrafish *prmt9* knockout model. Maternal zygotic mutants (*MZ*)*prmt9*<sup>-/-</sup> do not show a ciliopathy phenotype and the general brain development does not seem to be affected. However, after performing a social behavioral test in *MZprmt9*<sup>-/-</sup> adults, the velocity of mutants was significantly reduced compared to wildtype fish. Moreover, in contrast to the wild type animals, the mutants did not seem to become hectic in the immediate vicinity of their conspecifics, indicating a possible autistic phenotype.

In conclusion, in this first article I report the identification of a novel pathogenic mutations in the gene *PRMT9* which causes a syndromic form of autosomal recessive ID. This new finding can be an important

step towards a better understanding of the pathophysiology of neurodevelopmental disorders and could possibly aid the search for therapeutic targets and pathways.

### Personal contribution

For the following article, I performed the variant filtration of the WES data of the index family to identify the pathogenic mutation that causes the patient's phenotype and did segregation analysis for candidate genes. Moreover, I took care of the patient's fibroblasts and performed qPCR experiments and Western Blots to observe the effect of the identified mutation on the mRNA and protein level of patient's cells. I also performed the Western Blot experiments to confirm that mutated forms of PRMT9 are not able to bind to the splice factor SAP145 and performed immunostainings to measure the cilia length in fibroblasts of three affected individuals. With the help of Amélie Piton and Francesca Mattioli, I designed the RNA Sequencing experiment. RNA Sequencing data were pre-processed by Céline Keime and Damien Plassard (IGBMC) and then analyzed by myself using the annotation tool DAVID. In addition, I created a zebrafish knock down model, to prove the link between our gene of interest and the patient's phenotype. Finally, I wrote the article (Figure 12).



**Table 12. My personal contribution to project 1.** Icons were created with icons8.com.

#### 4.1.2 Introduction – Project 1

Intellectual disability (ID) is one of the major healthcare problems, affecting at least 1 % of the population worldwide.<sup>232</sup> ID is defined as a substantive limitation in intellectual functioning and adaptive behavior, with an age of onset before 18 years.<sup>20</sup> One can distinguish mild (IQ 50-70) and severe ID (IQ <50).

The major cause of ID is of genetic origin with pathogenic variants in a large number of genes as revealed by the efforts of many colleagues in large scale sequencing studies using either whole exome (WES) or whole genome (WGS) analysis.<sup>233</sup> A curated list of the monogenetic causes of ID (>700 genes) is included in the SysID database that classify each gene according to severity, penetrance additional associated clinical features and transmission mode.<sup>21</sup> Inherited in all forms (X-linked, autosomal dominant or recessive), researchers estimate that the number of autosomal recessive ID genes will especially continue to rise in the coming years.<sup>20</sup> Various functions have been implicated in ID including among others metabolism, transport, nervous system development, RNA metabolism and transcription. Remarkably, most enriched GO terms were the hedgehog pathway, glutamate receptor signaling, peroxisome, glycosylation and cilia.<sup>234</sup>

Clinically heterogeneous, it can occur either isolated or with additional anomalies in more complex syndromic forms.<sup>235</sup> Syndromic ID often appears with other abnormalities, like epilepsy or ASD.<sup>20</sup> Known syndromic disorders are, for example, Down syndrome, Fragile X syndrome or ciliopathies.<sup>20,235</sup> Ciliopathies are a group of syndromes associated with a dysfunction of the primary cilium or impaired primary cilium biogenesis. The primary cilium is an organelle that extends from the surface of almost all vertebrates and plays a critical role in developmental and homeostatic signaling pathways. In this case syndromic ID is often accompanied by additional symptoms like renal dysplasia, retinal degradation, limb abnormalities such as polydactyly, and a host of other symptoms.<sup>231</sup>

Due to this extreme genetic and clinical heterogeneity, it is often difficult to make a diagnosis.<sup>20</sup>

In this study, we report *PRMT9* as a novel gene for a recessive form of syndromic ID. *PRMT9* belongs to the family of protein arginine methyltransferases (PRMTs), a family of currently nine proteins whose main function is thought to be the methylation of arginine residues on histones or other proteins.<sup>167</sup> As a type II methyltransferase, *PRMT9* is able to generate monomethylarginines (MMAs) and symmetric dimethylarginines (SDMAs). Functionally, it is thought to play an important role in alternative splicing through its interactions with the splicing factor SAP145 and by modulating small nuclear ribonucleoprotein maturation.<sup>167</sup> *PRMT9* was already proposed as a potential candidate gene implicated in non-syndromic intellectual disability.<sup>170</sup> However, no functional studies were performed at that time and no additional families were available.

Starting from our index case with a ciliopathy like phenotype and thanks to world-wide collaborations, initiated using Decipher<sup>168</sup> and GeneMatcher<sup>169</sup>, we were able to assemble a cohort of 13 individuals from 9 families with 10 different *PRMT9* pathogenic loss of function variations. Beside mild to severe ID (12/13), the patients presented a global developmental delay (8/13), epilepsy (7/13), hypotonia (4/13), ASD (4/13), kidney anomalies (3/13), polydactyly (2/13) and minor facial anomalies. Functional studies implicating either patients' cells or zebrafish demonstrated the impact of the variations and revealed some of the pathway implicated in this disorder.

### 4.1.3 Results – Project 1

#### **Identification of Biallelic Loss of Function Variations in *PRMT9***

The index patient (A.II-1) is the first-born child of three children from a consanguineous marriage in a family of Algerian origin (Figure 12). Due to the presence of bilateral postaxial polydactyly, developmental delay and a suspected but not confirmed retinal dystrophy, a ciliopathy and more likely BBS, [MIM#20990]) was suspected (Table 13). Other clinical manifestations included poor muscle tone, bilateral cryptorchidism, spontaneously resolving vesicoureteral reflux and mild facial dysmorphology. His birth size and weight were within the normal range (2.7 kg, and 49 cm). With time, he demonstrated autistic behavior connected to major anxiety. At the age of seven, he had his first epileptic seizure, which has since been pharmacologically controlled. Cerebral magnetic resonance imaging (MRI) revealed the presence of bilateral periventricular nodular heterotopia (PVNH). At the last clinical examination at age 9, the patient had severe ID with a near complete lack of language. He had normal weight (+0.5 SD), height (-1.25 SD), and head circumference (+1.5 SD). Ophthalmological examination did not reveal any abnormalities.

**Table 13. Summary of the Major Clinical Features of Patients with biallelic Pathogenic Variants in *PRMT9***

Family	A	B	B	C	D	E	F	G	H	I			
Individual	A.II-1	B.II-1	B.II-6	B.II-7	C.II-1	C.II-2	D.II-1	E.II-3	E.II-5	F.II-3	G.II-1	H.II-1	I.II-2
<b>General information</b>													
Gender	M	M	F	M	F	M	M	M	M	M	n/a	M	F
Age (Y)	12	35	17	9	12	5	n/a	8	2	6 month	7	17	11
Ethnic origin	Algeria	Iran	Iran	Iran	n/a	n/a	Macedonia	Marocco	Marocco	Turkey	Syrian	Hungarian	Mother: Northern India/Europe/Jamaica/Cuba. Father: Ghana/Mauritania
Parental Consanguinity	+	+	+	+	-	-	-	+	+	+	+	n/a (adoption)	
Nucleotide change	545delT	565G>A	565G>A	565G>A	554G>A 1318C>T	554G>A 1318C>T	491delC	1918dup	1918dup	1076C>T	847_1330del	258_261delACTT	415G>T 554G>A
Amino acid change	Leu182Trpfs*1 0	Gly189Arg	Gly189Arg	Gly189Arg	Gly185Glu Gln440*	Gly185Glu Gln440*	Thr164Lysfs*28	Val640Glyfs*4	Val640Glyfs*4	Arg359*	Lys284Thrfs*3	Leu87Valfs*15	Glu139* Gly185Glu
Affected domain		Catalytic site (motif I)	Catalytic site (motif I)	Catalytic site (motif I)	Catalytic site (motif I)	Catalytic site (motif I)							Catalytic site (motif I)
<b>Neurodevelopmental Abnormalities</b>													
Intellectual Disability	severe	moderate-severe	moderate-severe	moderate-severe	mild	mild	mild	+	+	-	+	severe	moderate-severe
Developmental Delay	+	n/a	n/a	n/a	n/a	+	n/a	+	+	-	+	+	+
Speech impairment	+	n/a	n/a	n/a	n/a	+	n/a	+	+	n/a	+	+	+
Autism Spectrum Disorder	+	n/a	n/a	n/a	+	n/a	n/a	+	-	-	n/a	n/a	+
Behavioral abnormalities	major anxiety	n/a	n/a	n/a	n/a	n/a	hyperactivity	angers	-	n/ap	n/a	ADHD, major anxiety	Frequently brings objects and body parts to mouth.
Hypotonia	+	n/a	n/a	n/a	n/a	n/a	n/a	n/a	+	+	+	+	+
Seizures	+	-	-	-	+	+	+	+	+	-	-	+	-
CNS abnormalities on MRI	Periventricular heterotopia	n/a	n/a	n/a	n/a	White matter abnormalities	n/a	Wide pericerebral spaces	-	n/a	-	n/a	-
<b>Cranofacial Features</b>													
<b>Craniofacial features</b>													
	Flat facial profile			Micropthalmia			Otopostasis small ears	wide open eyes, low frontal hair line	Plagiocephaly, wide open eyes, low frontal hairline	prominent forehead, hypotelosim, broad flat nasal bridge, upslanting palpebral fissures, thin lips, blue sclerae, large, deep-set ears, short neck, flat occiput	flat facial profile, hypertelorism, epicanthal folds, broad nasal bridge, thin upper lip		bilateral strabismus, prominent forehead, broad nasal root, low hanging columella with thin alae nasi, prominent anthelix, high arch palate with dental crowding, everted lower lip, elongated neck
<b>Skeletal-Limb Features</b>													
6th finger or toe abnormalities	polydactyly	-	-	polydactyly	-	-	-	-	-	-	-	"curled toes"	-
Others								hyperlaxity	Hyperlaxity	skeletal dysplasia, disproportional short stature, slight platyspondyly, cone-shaped ventral ribs, cone-shaped distal ulna, shortening of radius, humerus and femur, shortening of phalanges (middle and distal phalanx)			
<b>Additional symptoms</b>													
Kidney abnormalities	Spontaneously resolving pelviureteric junction obstruction	n/a	n/a	n/a	n/a	n/a	Prenatal hydronephrosis, BL VUR grade IV	-	-	-	n/a	pending ultrasound, 2 urinary tract infections	n/a
Cryptorchidism	+	-	-	-	-	+	-	+	-	-	-	-	-
Supernumerary nipple	-	-	-	-	-	-	+	-	+	-	-	+	-
Additional symptoms							Ventricular septal defect hyperprolinemia	constipation			Ataxia and tremor, repeated ENT infections, ventilation tubes, tonsillectomy, hairy legs and arms, 4-5 Cafe au lait patches on her torso	requires NGT feeding, Ventricular septal defect, eczema, recurrent infections	Chest: Flat with delayed breast development, delayed puberty Tanner stage: 2.

After negative screening using a panel based approach of known BBS genes and related phenotypes<sup>236</sup>, WES was performed. A total of 56,911 single nucleotide substitution and insertion-deletion variants were identified in the protein-coding sequences (supplemental Figure S1.1). Thanks to the classical filtering strategy including functional criteria, frequency in population-based databases and cosegregation analysis (see methods), the resulting 13 homozygous variants were further filtered out based on their sequencing quality, gene function and implication in other diseases. This led to the shortlisting of 2 potentially pathogenic variants: a homozygous nonsense variation (c.511G>T, p.Gly171\*) in the gene *PRSS33* (NM\_152891.2) covered by 2 reads and a homozygous single nucleotide deletion (c.545delT, p.Leu182Trpfs\*10) leading to a frameshift and premature termination codon (PTC) in the gene *PRMT9* (NM\_138364.2) covered by 97 reads. Moreover, we screened WES data for variations in known PVNH and ID genes, but were able to exclude all potential candidates by Sanger sequencing (supplemental Figure S1.1+S1.2). As Sanger sequencing proved that the variant found in *PRSS33* was a false positive call, we subsequently focused on the candidate gene *PRMT9*.

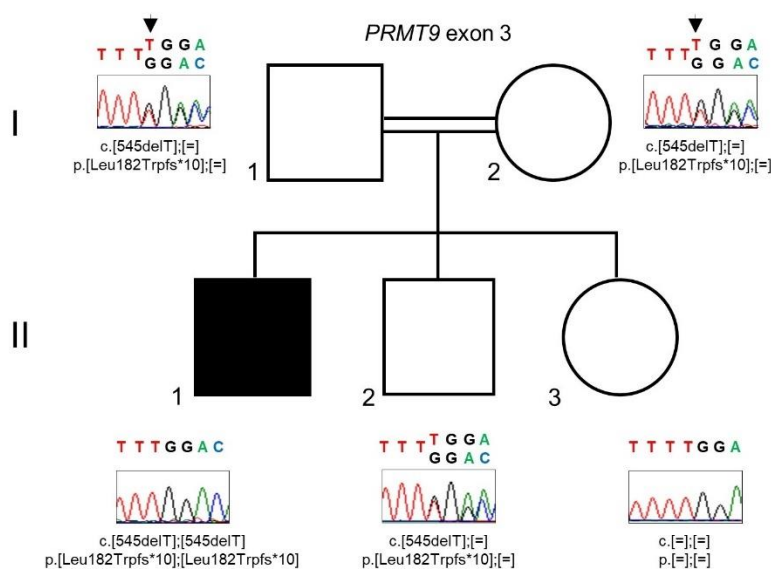


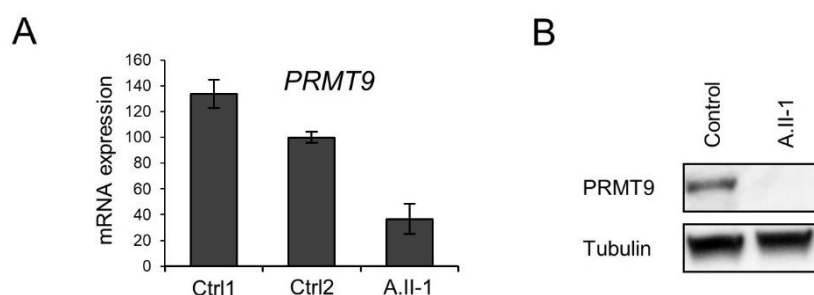
Figure 12. Identification of a loss of function mutation in the gene *PRMT9* in the index family A.

### Effect of the homozygous frameshift variation on *PRMT9* expression

In order to investigate the effect of the identified variation (p.Leu182Trpfs\*) in *PRMT9*, we analyzed the mRNA and protein levels in the patient's skin fibroblasts. In three independent experiments, the expression of *PRMT9* mRNA was decreased in patient's cells by ~60 % compared to the control cells (Figure 13A), suggesting an involvement of the nonsense-mediated mRNA decay (NMD) pathway. Moreover, we examined the effect of the premature stop codon at the protein level of *PRMT9*. For this, proteins from patient's fibroblasts and control cells were extracted and analyzed by Western blot using an antibody specific to the N-terminal of *PRMT9*. However, we observed neither the full-length



protein (Figure 13B) nor a truncated form (data not shown) of PRMT9 in patient's cells, suggesting that the shortened protein is unstable and is likely degraded by the ubiquitin-proteasome system.

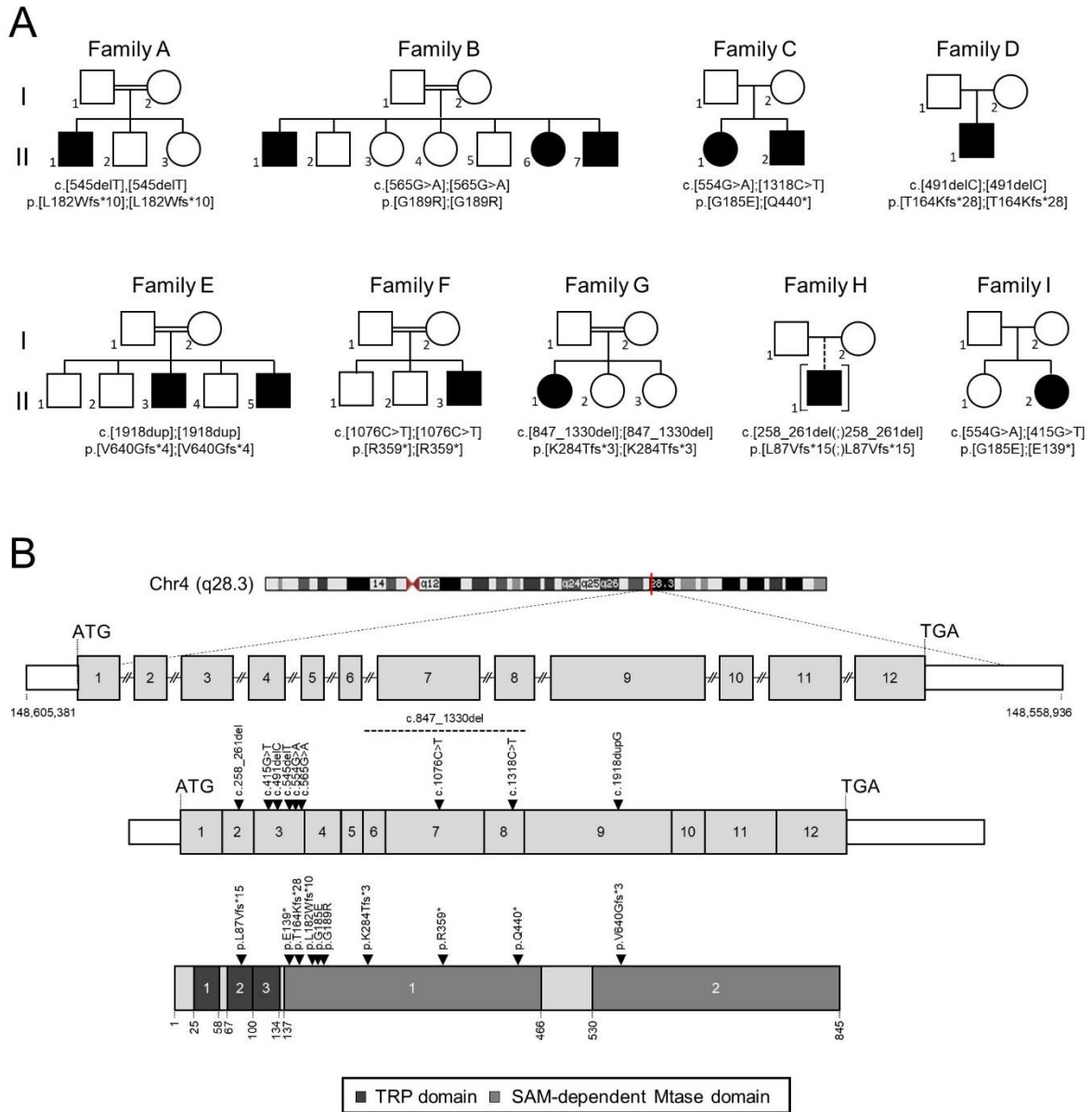


**Figure 13. Effect of the *PRMT9* variation p.Leu182Trpfs\* on the mRNA and protein expression.**

**(A)** Real-time qPCR analyses of the expression of *PRMT9*, normalized to GAPDH and β-Actin in control and patient fibroblasts. Error bars represent standard deviation of technical triplicates. **(B)** Extracted proteins from fibroblasts were analyzed by Western blotting. β-Tubulin serves as a loading control.

### Recruitment of further individuals with *PRMT9* variations

Using Decipher<sup>168</sup> and GeneMatcher<sup>169</sup>, we were able to recruit 8 additional families, of which 5 were from consanguineous union from, with affected family members (from 1 to 3 per family) carrying loss of function (LoF) variations in *PRMT9* (Figure 14A and Table 13). *PRMT9* includes 12 exons and is located on chromosome 4 (4q31.23) and has a genomic size of 46.4 kb. The mRNA transcript is 2 kb long and codes for a protein of 845 AA (94.5 kDa) (Figure 14B). In total, 13 individuals carried biallelic pathogenic variations in *PRMT9*. Among the 10 different variations, one can observe 4 frameshifting indels, 3 nonsenses, 2 missenses and 1 large deletions (exons 6 to 8). Interestingly, the 2 missenses are located within 2 of the 3 highly conserved glycine in the core motif (Motif I) defining the AdoMet binding pocket, one of the critical catalytic site of PRMTs<sup>237</sup> (supplemental Figure S1.3).



**Figure 14. *PRMT9* mutations found in nine families.**

**(A)** Pedigrees of nine families with thirteen affected individuals. Pedigrees have been drawn according to published standards.<sup>238</sup> We are currently collecting all co-segregation data. **(B)** Schematic demonstration of the gene *PRMT9*, mRNA and protein with the different protein domains. Mutations are indicated in different colors.

### Clinical features of individuals carrying *PRMT9* variations

Comparing the phenotypic characteristics of the thirteen affected individuals (Table 13 and 14), we noticed the presence of mild to severe ID in all individuals, except a six-month-old Syrian boy (G.II-1). With regard to the age of the young proband, the development of his intellectual abilities and other symptoms remains to be seen in the future. Moreover, clinical examinations revealed that 7 of 13 individuals developed epilepsy. A brain MRI was conducted in 4 of the 7 affected individuals showing abnormalities in two of them (A.II-1 Periventricular heterotopia and C.II-2 white matter abnormalities). A global developmental delay was observed in 39 % (5/13), as well as a poor expressive language in 54

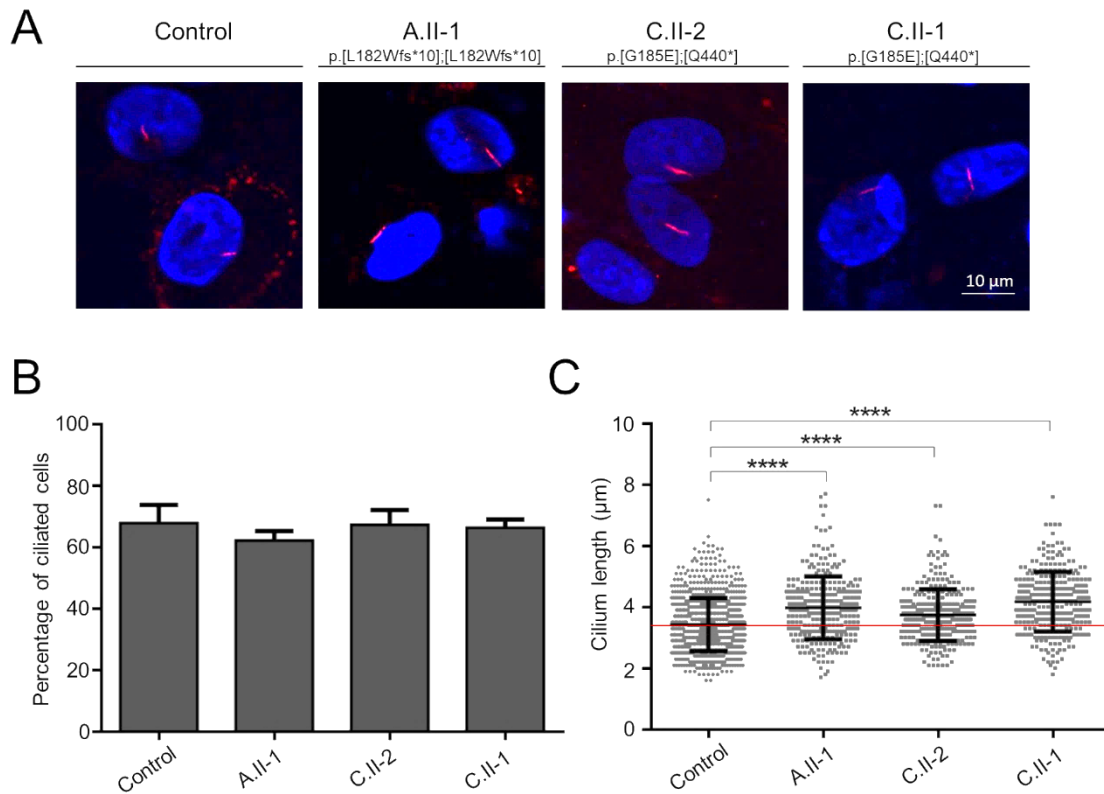
% (7/13). ASD was diagnosed in 31 % (4/13) of affected individuals. Moreover, many individuals presented congenital malformations, including hypotonia (5/13), cryptorchidism (3/13) and urogenital malformations (3/13). In 7 of 13 individuals facial anomalies were seen, including a flat facial profile, hypertelorism and a prominent forehead. Skeletal anomalies were noticed in 6 of 13 individuals with 15 % (2/13) presenting polydactyly and 15 % (2/13) hyperlaxity.

**Table 14. Distribution of Clinical characteristics with the PRMT9 cohort.**

	Percentage	Number of affected individuals
<b>Neurological</b>		
Intellectual disability, mild-severe	92 %	12/13
White/Grey matter anomalies	15 %	2/13
Epilepsy	54 %	7/13
Autistic behavior	31 %	4/13
Hypotonia	39 %	5/13
<b>Development</b>		
Global developmental delay	39 %	5/13
Delayed speech and language development	54 %	7/13
<b>Congenital</b>		
Postaxial polydactyly	15 %	2/13
Cryptorchidism	23 %	3/13
<b>Kidney defects</b>	23 %	3/13
<b>Facial dysmorphism</b>	54 %	7/13

### Primary Cilium Length is Increased in Affected Fibroblasts Compared to Control Cells

Given the initial suspicion of a ciliopathy phenotype in the index patient (A.II-1) including polydactyly (2 patients) and other skeletal malformations (6 patients), which can often be caused by defective Shh signaling at the primary cilium, cryptorchidism as well as kidney anomalies in three individuals, we examined whether the cilia of patients could be affected. To address this we compared the patients' cells (skin fibroblast from individuals A.II-1, C.II-1 and C.II-2) under ciliary conditions (-FCS) with control cells. Counting the number of ciliated cells did not show any difference (Figure 15A and B). Nevertheless, measuring the primary cilia length showed a significantly longer cilia in patients' cells than in the control (Figure 15C). These results indicate a possible effect on the ciliary biogenesis or function. Interestingly, these observations were similar using either a fully depleted PRMT9 cell (A.II-1) or a half reduced amount of PRMT9 and a defective PRMT9 protein (C.II-1 and C.II-2).



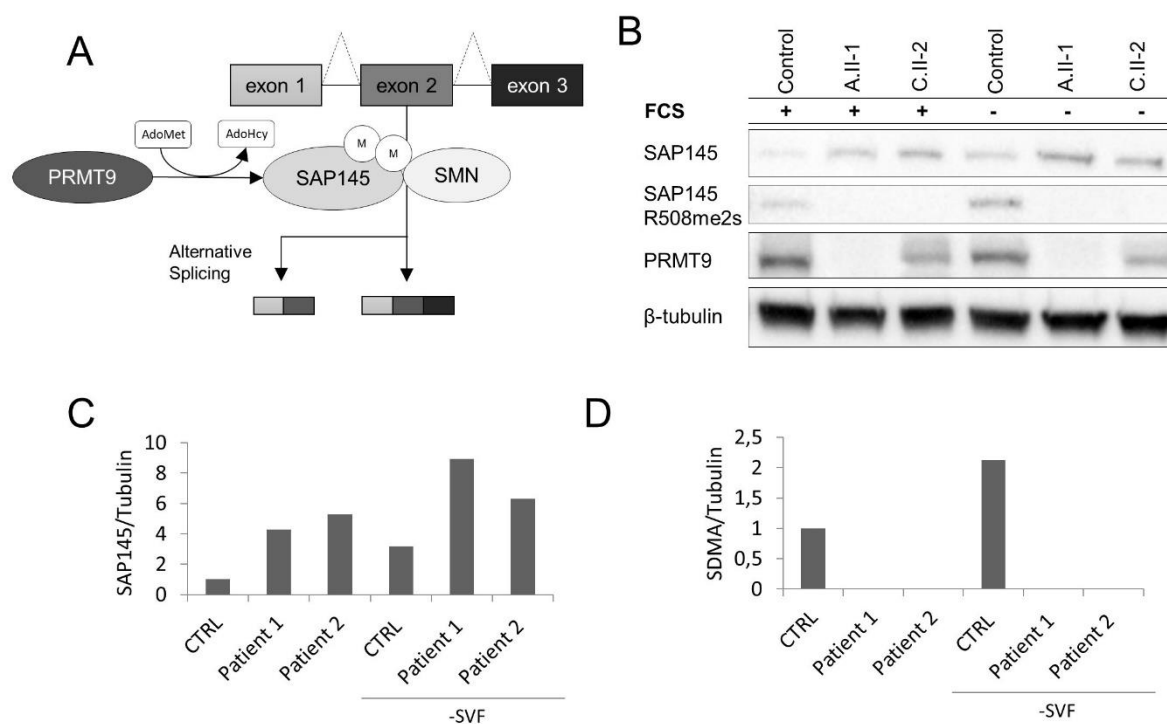
**Figure 15. Mutations in the gene *PRMT9* affect the cilia length in patient fibroblasts.**

**(A)** To induce primary cilium formation, the cells were deprived of serum by growth for 48 h in DMEM with 1 % Pen/Strep. Fibroblasts from control and patient cells were stained with an antibody against acetylated  $\alpha$ -Tubulin (cilia, red). Nuclei were stained with Hoechst (blue). **(B)** The proportion of ciliated cells is unaltered in patients' cells compared to control cells. The control Ctrl is the mean of ctrl1, 2 and 3 and error bars represent standard deviation (sd). **(C)** Patients' cells present longer cilia. Data are presented as a scatter plot with a line indicating the mean value and error bars indicating the standard deviation (sd). The control Ctrl is the mean of ctrl1, 2 and 3. Statistical significance was determined using the unpaired, nonparametric Mann-Whitney Test, ns = non-significant. \*  $p < 0.05$ . \*\*  $p < 0.01$ . \*\*\*  $p < 0.001$ . \*\*\*\*  $p < 0.0001$ . Significance is determined relative to control cells.  $n=300$ .

### **PRMT9 Fails to Methylate the Splicing Factor SAP145 in Affected Fibroblasts and Under Ciliated Conditions**

PRMT9 is known to interact with the splice factor SAP145<sup>167</sup> and especially methylate SAP145<sup>R508me2s</sup> in order to stimulate the SAP145 interaction with SMN (Survival of motor neuron), a protein important for the assembly of small nuclear ribonucleoprotein particles essential for pre-mRNA splicing (Figure 16A).<sup>167</sup> To examine the effect of the *PRMT9* LoF, a Western blot was performed to determine if PRMT9 was still able to methylate SAP145, by using an antibody specific for symmetrically dimethylated Arginine 508 of SAP145 (SDMA)<sup>167</sup>. While SAP145 was detected in both control and patient cells (A-II.1 and C-II.2) cells, SDMA was only detected in the control group, showing that the *PRMT9* variants identified in family A and C prevent the methylation of the splice factor (Figure 16B). Since we have shown that patients' cells have an increased primary cilium length, we additionally investigated by western blot, if the methylation of SAP145 plays a predominant role in ciliated conditions. Interestingly, SAP145 as well as the dimethylated form of SAP145 (SDMA) were upregulated in starved

conditions (-FCS), highlighting a potential role of PRMT9 during ciliogenesis (Figure 16C-D). However, since the experiment was only performed once, it needs to be repeated to confirm this hypothesis.



**Figure 16. In affected individuals, PRMT9 fails to methylate the splicing factor SAP145.**

(A) PRMT9 is known to symmetrically dimethylate Arginine 508 of SAP145. Due to the interaction with the splicing factor SAP145, PRMT9 was suggested to regulate alternative splicing.<sup>167</sup> (B) Control cells and patients fibroblasts (individuals A.II-1 and C.II-2) were cultured in normal conditions (+FCS) and ciliated conditions (-FCS). To induce primary cilium formation, the cells were deprived of serum by growth for 48h in DMEM with 1 % Pen/Strep. Extracted proteins were analyzed by Western blotting.  $\beta$ -Tubulin serves as a loading control. (C-D) Quantitative analysis of the western blot (n=1).

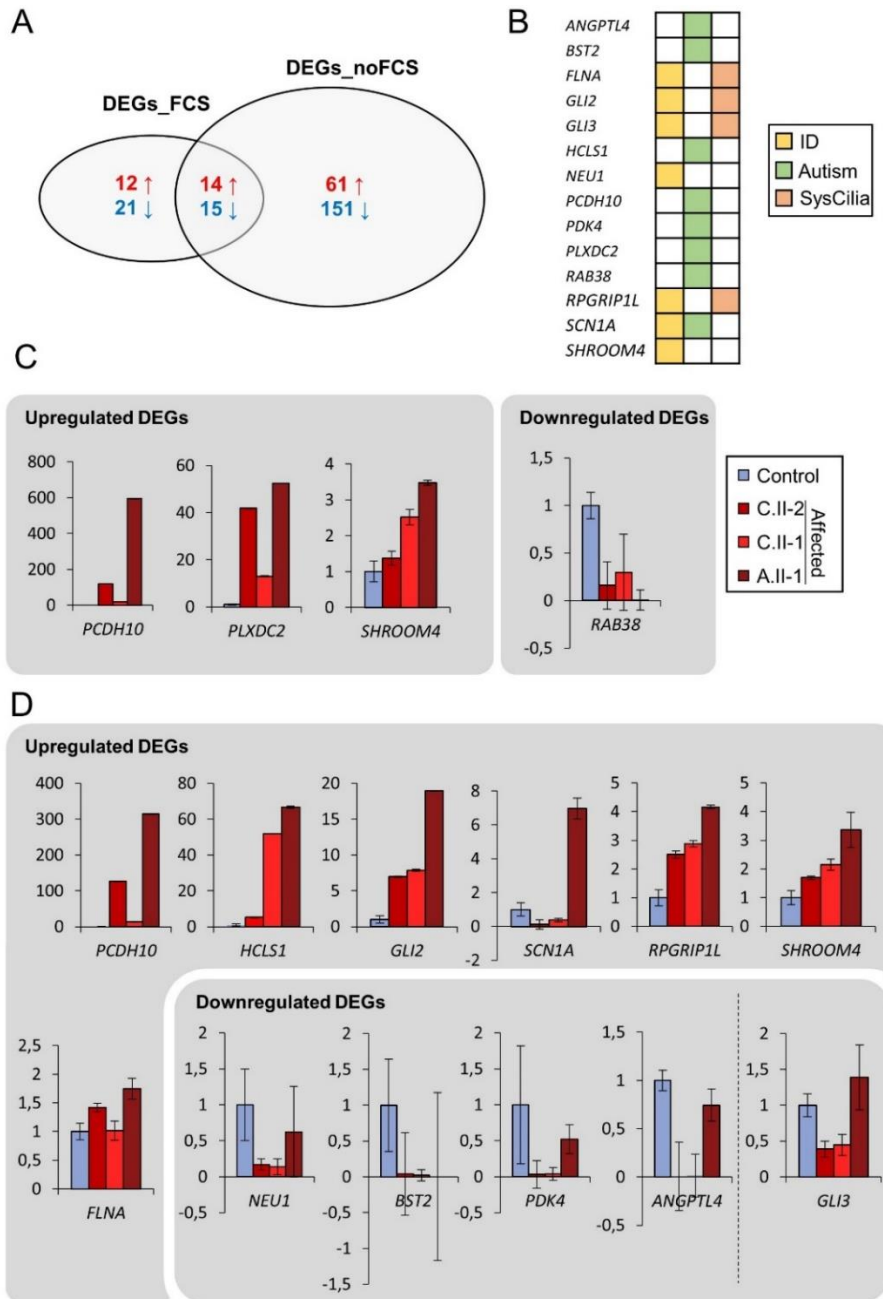
### Biallelic *PRMT9* Variations Affect the Expression of Genes Associated with ID, ASD and Cilia

In order to identify the molecular mechanisms underlying the patients' phenotype, we examined the RNA differential expression with a specific focus on genes associated with ID<sup>20,239</sup>, ASD<sup>240</sup> and cilia function/biogenesis<sup>241</sup>. The transcriptomic study (RNA sequencing) was performed on the same 3 samples (A.II-1, C.II-1 and II-2) under either rich medium conditions (+FCS) or under ciliated conditions (-FCS). To detect the effect of *PRMT9* variations, we focused only on shared differently expressed genes (DEGs) of all three affected individuals. Focusing on an at least a log<sub>2</sub> fold change (FC) of >1/<-1 and an adjusted p-value of <0.05, 62 DEGs were found in cells cultured with FCS (supplemental Table S1.2) of which 26 were upregulated and 36 downregulated (Figure 17A and supplemental Table S1.2). Considering the cells under ciliary conditions (-FCS), 241 DEGs could be identified of which 75 were upregulated and 166 downregulated (Figure 17A and supplemental Table S1.4).

**Table 15 Enrichment in Genes Involved or Potentially Involved in ID, ASD or cilia function/biogenesis.** A total of 20,000 protein- coding genes were taken as population size for the hypergeometric tests. Statistical significance of the overlap between two groups was tested using the online tool “[http://nemates.org/MA/progs/overlap\\_stats.html](http://nemates.org/MA/progs/overlap_stats.html)”

	DEGs	Enrichment (No. Of Genes in Overlap)	p Value	Genes
<b>DEGs (FCS)</b>				
ID (n=719)	(n=62)	0.9 (2)	p < 0.386	<i>SHROOM4, SATB2</i>
DE in ASD From AutismKb (n=1660)	(n=62)	1.6 (8)	p < 0.140	<i>PCDH10, HCLS1, PLXDC2, RTN1, TIPARP, CD44, DLX1, RAB38</i>
SysCilia (n=303)	(n=62)	0.0 (0)	p < 0.388	
<b>DEGs (noFCS)</b>				
ID (n=719)	(n=241)	1.8 (16)	p < 0.014	<i>SHANK2, RAB39B, GLI2, PYCR1, PRPS1, SHROOM4, SLC1A4, FLNA, RPGRIP1L, GLI3, SYNGAP1, SPATA13, NEU1, GRM1, NEDD4L, SCN1A</i>
DE in ASD From AutismKb (n=1660)	(n=241)	2.1 (43)	p < 1.346e-06	<i>SLC16A9, HCLS1, PCDH10, RTN1, PCDH7, RAB39B, MYOM2, BCAT1, ANXA2, C7orf50, PPIC, CD44, NBPF10, AHRR, SH2B2, ZNF385A, RALGPS2, GYPC, LIFR, STOM, SPATA13, H1FO, ETS2, ANKRD9, ANGPTL4, CLEC2B, ADORA2B, ITPKA, IFITM1, RHBDL3, ABCA1, PLIN2, SOD2, ATP8B4, RAB38, PDK4, NEDD4L, CMPK2, VGF, SCN1A, RSAD2, PPARGC1A, BST2</i>
SysCilia (n=303)	(n=241)	1.1 (4)	p < 0.497	<i>GLI2, FLNA, RPGRIP1L, GLI3</i>

Among the significant DEGs identified in rich conditions (+FCS), two genes, *SHROOM4* and *SATB2*, are known to be associated with ID and eight genes, *PCDH10, HCLS1, PLXDC2, RTN1, TIPARP, CD44, DLX1* and *RAB38* with ASD (Table 15 and supplemental Table S1.2). Performing a RT-qPCR, we confirmed the upregulation of *PCDH10, PLXDC2* and *SHROOM4* and the downregulation of *RAB38* (Figure 17B-C). Enriched gene ontology (GO) categories were determined using the functional annotation online tool DAVID (6.8)<sup>242,243</sup>. Affected biological processes (BPs) are e.g. cell adhesion (GO: 0007155) or the negative regulation of transcription from RNA polymerase II promoter (GO: 0000122) (Table S1.3). Under ciliated conditions (-FCS), the number of DEGs increased in affected fibroblasts with a significant enrichment of DEGs involved in ID (p-value <0.014) and/or ASD (p-value <1.346E<sup>-6</sup>) (Table 15). RNA sequencing data also revealed the differential expression of *GLI2, GLI3, FLNA* and *RPGRIP1L*, four genes listed in the SYSCILIA gold standard (SCGSv1)<sup>241</sup>, a database of known ciliary components, strengthening the hypothesis that PRMT9 might have a direct or secondary impact on ciliogenesis or cilia function (Table 15 and supplemental Table S1.4). RT-qPCR was performed to validate the RNA Sequencing results a selection of upregulated genes (*PCDH10, HCLS1, RPGRIP1L, SHROOM4* and *FLNA*), as well as downregulated genes (Table 17).



**Figure 17. Multiple genes associated with ID, autism or cilia biogenesis/ function are dysregulated in individuals carrying biallelic PRMT9 mutations.**

**(A)** Prior to RNA extraction and sequencing, control cells and patients' fibroblasts (individuals AII-1, CII-15 and CII-2) were cultured in normal conditions (+FCS) and ciliated conditions (-FCS). To induce primary cilium formation, the cells were deprived of serum by growth for 48h in DMEM with 1 % Pen/Strep. Considering only genes with a log2 fold change of <-1/>1 and an adjusted p-value <0,05, 62 DEGs were identified in all three patients in normal cell culture conditions (+FCS) and 241 DEGs were found in ciliated conditions (-FCS) highlighting a potential role of PRMT9 during ciliogenesis. **(B)** The dysregulation of 14 of 62 genes associated with ID, autism or cilia biogenesis/function was further validated by qPCR. **(C)** Real-time qPCR analyses of the expression of upregulate and downregulated genes in patients' fibroblasts culture in normal conditions (+FCS). **(D)** Real-time qPCR analyses of the expression of upregulate and downregulated genes in patients' fibroblasts culture in ciliated conditions (-FCS). Error bars represent standard deviation of biological triplicates.

(*NEU1*, *BST2*, *PKD4* and *ANGPTL4*). Unlike the RNA Sequencing results, *GLI3* is downregulated only in Enriched BPs are e.g. actin cytoskeleton organization (GO: 0030036), negative regulation of transcription, DNA-templated (GO: 0045892), regulation of synaptic plasticity (GO:0048167), cell adhesion (GO:0007155), small GTPase mediated signal transduction (GO:0007264), positive regulation of transcription from RNA polymerase II promoter (GO:0045944) (Table S1.5).

### **Biallelic *PRMT9* Variations Affect the Alternative Splicing of some Genes**

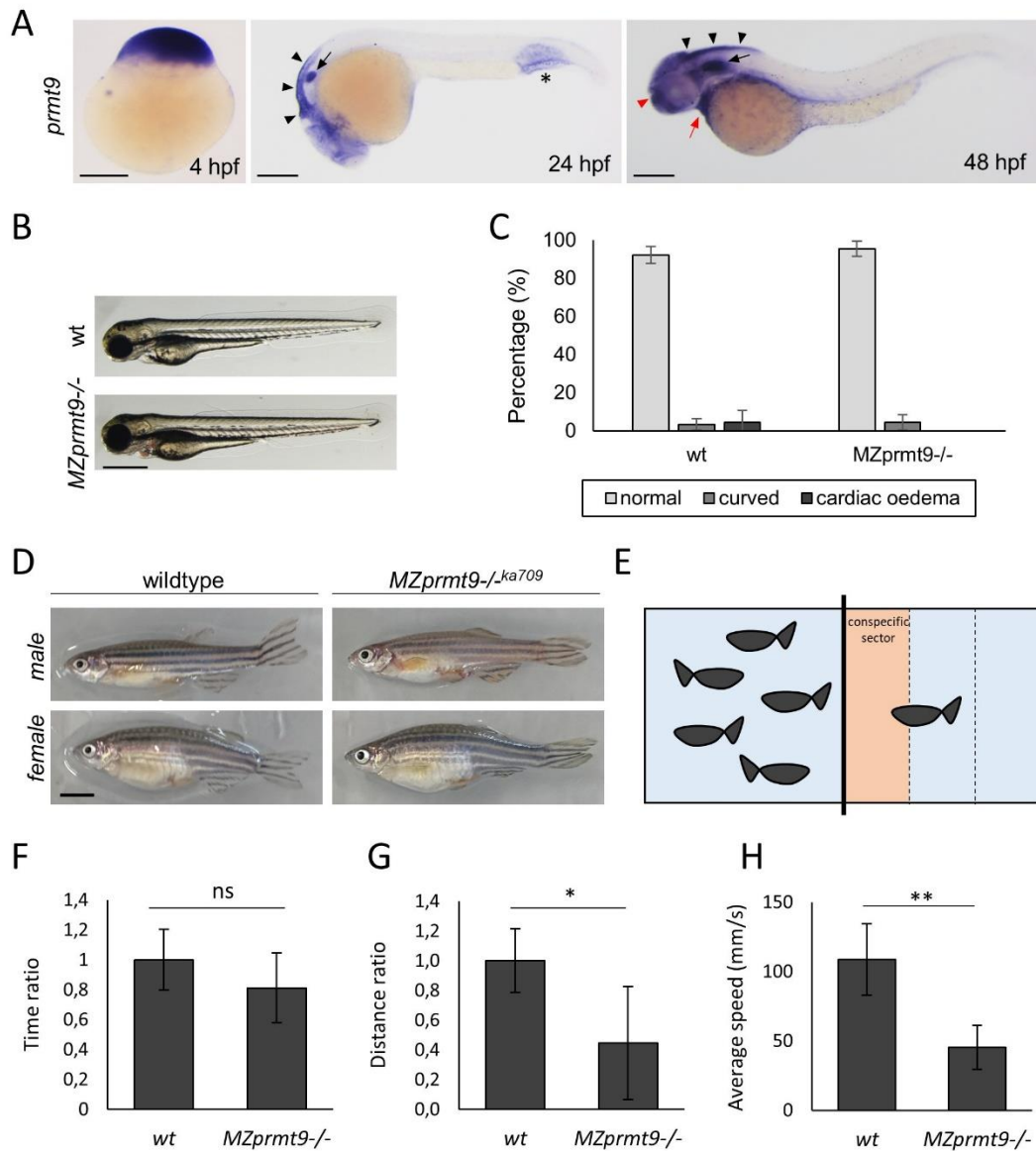
Since, *PRMT9* was shown to be unable to methylate the splice factor *SAP145* in patients' fibroblasts (A-II.1 and C-II.2), we thought to explore the effect on alternative splicing. Due to the increase of methylated *SAP145* in ciliated conditions, we assumed that this effect might be especially remarkable after serum starvation. Using Leafcutter (filter: p.adj.<0.01; logef >0.6 or <-0.6;  $\Delta\Psi$ <-0.5 or  $\Delta\Psi$ >0.5) and JunctionSeq (filter: log2FC >3 or <-3; p.adj. <0.01; dispersion < 0.05), we indeed observed more common alternative spliced genes in ciliated fibroblasts of all three patients (supplemental Figure S1.4). In patients' cells cultivated with FCS, 16 alternative spliced genes (13 detected by JunctionSeq and 3 by Leafcutter) were detected. Only one of the genes was previously associated with ID and epilepsy: *GJC2* (MIM: 608803). Moreover, the alternative spliced gene *NME1*, is thought to play a role during neuronal cell fate determination<sup>244</sup> (supplemental Table S1.6). In ciliated conditions (-FCS), alternative splice events were observed in 19 genes by JunctionSeq and 10 genes by Leafcutter, with *SCRG1* detected by both tools. Among those 18 genes, three have previously been associated with neurogenesis and neuronal cell fate determination: *PDGFRB*<sup>245</sup>, *SCRG1*<sup>246</sup>, *NME1*<sup>244</sup> (supplemental Table S1.7).

### **Zebrafish *MZprmt9*<sup>/ka709</sup> Mutants Display Behavior Anomalies**

Zebrafish has become with time a precious model organism for human Mendelian diseases, such as ciliopathies<sup>247</sup> and neurodevelopmental disorders e.g. ASD or epilepsy.<sup>248</sup> The human *PRMT9* gene has a single orthologue (ENSDARG00000036755) in the zebrafish genome, located on the reverse strand of chromosome 1 with two predicted isoforms (859aa and 876aa and 5aa).<sup>249</sup> With a sequence identity of 61 % to the human orthologue, in this study we considered only the two longer isoforms. To investigate the effect of *PRMT9* LoF variations on the embryonic development and adult behavior of zebrafish, a CRISPR/Cas9 directed gene knockout was performed (supplemental Figure S1.5). The guide RNA binds a sequence within exon 4 of the *prmt9* gene and led to an insertion of four base pairs resulting in a frameshift and a premature stop codon. This results in a putative 269 amino acids (aa) *prmt9* protein instead of 876 aa. Real-time qPCR analysis of the expression of *PRMT9*, normalized to  $\beta$ -actin in wildtype and *prmt9*<sup>-/-</sup> mutants, revealed a drastic drop of *prmt9* mRNA expression, indicating thus NMD. *In situ* hybridization for *prmt9* after 4 hours post fertilization (hpf) showed that *prmt9* is



maternally expressed. After 24 hpf an increased *prmt9* expression was detected in the brain (cerebellum and rhombencephalon), the otic capsule, and the blood island and after 48 hpf mRNA expression was additionally observed in the heart, epiphysis, and telencephalon (Figure 18A).



**Figure 18. *MZprmt9<sup>-/-</sup>* mutants do not display a typical ciliopathy phenotype, but show behavioral abnormalities.**

(A) Whole mount in situ hybridization analysis with *prmt9* antisense probe revealed a maternal contribution of *prmt9* mRNA (4 hpf). At 24 hpf high expression levels of *prmt9* were detected in the hindbrain (black arrowhead), the otic capsule (black arrow) and the blood island (black asterisk). At 48 hpf, additional staining was observed in the forebrain (read arrowhead) and the heart (red arrow). Scale bars 250  $\mu$ m. (B) Representative images of a 3 dpf old wildtype and *MZprmt9<sup>-/-</sup>* mutant zebrafish embryo. Scale bar 500  $\mu$ m. (C) Zebrafish embryos did not display obvious malformations. (D) Alike mutant embryos, *MZprmt9<sup>-/-</sup>* adults (12 mpf) did not display obvious body malformations. Scale bar 500  $\mu$ m. (E) Schematic representation of the social behavior test. (F) Ratio of the time spend in the conspecific sector. (G) The ratio of the distance traveled in the conspecific sector is significantly reduced in *MZprmt9<sup>-/-</sup>* mutants compared to wildtype fish. (H) The average speed of *MZprmt9<sup>-/-</sup>* mutants was significantly decreased in all three compartments. n=6/group.

To exclude the maternal contribution, maternal zygotic (MZ) *prmt9*<sup>-/-</sup> mutants were used to examine the embryonic development of *prmt9* deficient fish. Given a potential implication of an affected primary cilium in patients' fibroblasts, we examined 3 days post fertilization (dpf) old embryos for typical features observed in previously described ciliopathy zebrafish models.<sup>247</sup> However, neither a curved body axis, hydrocephalus, nor kidney cysts could be observed in *prmt9* mutants (Figure 18B-C). Since brain malformations (3/13) and ASD (4/13), a feature often accompanied by brain abnormalities in both fish and humans<sup>250</sup>, have been described in our probands, we next examined possible morphological brain changes in *prmt9* morphants. For this, *in situ* hybridization targeting the brain markers *krox20* and *msxc* was performed on 24 hpf embryos but no obvious malformations of the brain structures were observed in *prmt9* deficient embryos (supplemental Figure S1.6A). Via immunostaining, we next examined the axon tracts in the forebrain of 35 hpf embryos using an antibody against acetylated Tubulin, since RNA sequencing in affected human fibroblasts revealed an abnormal high expression of *PCDH10*, a gene that has been shown to play a role in axon outgrowth in the forebrain of mice. However, both wildtype and mutant *MZprmt9*<sup>-/-</sup> embryos presented intact axonal bundles, representing the main white matter tracts, in the forebrain (supplemental Figure S1.6B).

In addition, a morpholino targeting the start site of *prmt9* to prevent its translation was used to examine the function of *prmt9*. The efficiency of the morpholino was validated by co-injecting a *prmt9* N-terminal sequence fused to GFP and the *prmt9* morpholino. *prmt9-mo* efficiently binds to *prmt9*<sup>5'UTR-ex1</sup>-gfp and prevents the expression of the construct (supplemental Figure S1.7). In contrast to *prmt9* knockout mutants, 24 % of *prmt9* morphants presented a curved body axis (n=208) whereas only 3 % of 5 bp mismatched morpholino injected controls (n=184) and wildtype fish (n=126) displayed an abnormal body axis (supplemental Figure S1.8A-B). Further ciliopathy signs were not observed. Moreover, the hindbrain ventricle opening, which usually takes place before 24 hpf was impaired in *prmt9-mo* injected embryos. Performing an *in-situ* hybridization in 24 hpf old embryos against *mscx* and *krox20* both brain markers indicated that the brain ventricle is still closed in 25 % (*krox20* n=20, arrowhead) to 47 % (*mscx* n=15, arrow) of *prmt9-mo* injected embryos whereas the hindbrain ventricle completely open in all control injected embryos (*ctrl-mo*, *mscx* n=13, *krox20* n=14) (supplemental Figure S1.8C). Alike in *MZprmt9*<sup>-/-</sup> mutants the axon bundles of *prmt9-mo* injected embryos were not affected (supplemental Figure S1.8D). However, only one single morpholino was used for this study. To exclude a potential off-target effect of *prmt9-mo* further experiments are required, e.g. by using a second morpholino or performing a rescue experiment by co-injecting of the *prmt9* mRNA.

Also, in the research field of ASD, the popularity of zebrafish models is rising steadily.<sup>248</sup> A frequently conducted assay to investigate ASD zebrafish models is to test the social preference of adult animals.

For this one makes use of the swarm behavior of the species. The spatial segregation of a single fish from its group, in a tank that is divided only by a transparent disc, in general leads to the fact that the single fish stays most of the time close to its conspecific group.<sup>145</sup> ASD zebrafish models, however, do not show this strong social behavior and swim evenly throughout their whole compartment.<sup>251</sup> To test the behavior of *MZprmt9<sup>-/-</sup>* mutants, we made use of a common breeding tank divided into two compartments with a clear barrier and placed a group of five conspecific fish on one side and a single mutant or wildtype fish on the other side (Figure 18E). While we did not observe any difference in the time spend in the conspecific sector between wild type and *MZprmt9<sup>-/-</sup>* mutants, the distance traveled in the conspecific sector of the two groups differed significantly (Figure 18F-G). Mutants moved significantly slower in all three sectors (Figure 18H) and moved less hectically in the immediate vicinity of the conspecific group (Movie S1).

#### 4.1.4 Discussion – Project 1

In this study, we report 9 families with 13 affected individuals presenting a syndromic form of intellectual disability associated with epilepsy, autism, a global developmental delay, an impaired speech development, various skeletal anomalies (including polydactyly) and hypotonia.

As a proven strategy for gene identification for such homozygous condition (ID), whole exome sequencing has been applied worldwide<sup>252</sup> to several families and thanks to data sharing via GeneMatcher<sup>169</sup> or DECIPHER<sup>168</sup>, we were able to assemble this cohort. Among those families, 5 were of consanguineous unions and all affected individuals carry biallelic inherited pathogenic variations in the *PRMT9* gene. Since all known *PRMT9* variations are either leading to a premature stop codon or/and to a missense variation in the catalytic site (Motif I the S-adenosyl methionine co-factor binding site), we assume that only LoF variations cause the phenotype.

Interestingly in 2011, Najmabadi and colleagues (Family B)<sup>170</sup> already proposed *PRMT9* (aka *PRMT10* at this time) as a candidate ID gene but as an isolated form. At that time, the missense variation (p.Gly189Arg) could not be further studied and since then no other family was reported. Reinvestigation of the family herein (PhD) shed light on additional clinical manifestations such as polydactyly in one of the affected family members.

*PRMT9* encodes for one of nine protein arginine methyltransferases (PRMTs), a protein family that can be classified into three different groups according to their methylation products. The majority of PRMTs (*PRMT1*, *PRMT2*, *PRMT3*, *PRMT4*, *PRMT6*, and *PRMT8*) are able to generate monomethylarginines (MMAs) and asymmetrical dimethylarginines (ADMAs) on their targets. Unlike these type I PRMTs, *PRMT9* and *PRMT5*, the only two members of type II PRMTs, are capable of forming MMAs and symmetric dimethylarginines (SDMAs). The sole type III member *PRMT7* only creates a MMA mark.<sup>167</sup> Since PRMTs have a wide spectrum of different targets, like transcription factors, ion channels, splicing factors, scaffolding proteins or transport proteins, it's not surprising that they have been shown to take part in many different processes. For example, they have an influence on RNA splicing, tumor suppression, DNA repair and hormone receptor signaling.<sup>230</sup> Franic and colleagues revealed a strong association between *PRMT9* and intelligence.<sup>253</sup> Interestingly, variations in *PRMT7*, the gene most closely related to *PRMT9*, were identified in six patients with mild ID, obesity and shortening of the digits.<sup>254,255</sup> Since *PRMT9* was identified only 12 years ago by homology to *PRMT7*<sup>256</sup>, little is known about the protein yet. As mentioned and warned by Yang *et al*, *PRMT9* is the official gene name for this gene located on chromosome 4 (4q31.23)<sup>167</sup>. It was previously referred as *PRMT10* (as used in 2011<sup>257</sup>) and wrongly used for naming another gene on chromosome 2p16 *FBXO11*. This confusion although clearly stated by the HUGO Gene Nomenclature Committee HGNC can lead to unclear description of gene function as in Jansen *et al*<sup>258</sup>.

Considering the presence of postaxial polydactyly and ID in the index family, signs also present in ciliopathies, an involvement of PRMT9 in the structure or function of the cilia was initially suspected. Indeed, we were able to show that primary cilia are longer in individuals with *PRMT9* LoF variations. However, a connection between the cilia and the protein PRMT9 is not known yet. But in two independent proteomic analysis, in which primary cilia have been isolated from mice kidney cells or swine choroid plexus epithelial cells by using a calcium-shock method, *PRMT5* has been identified as a candidate ciliary gene by mass spectrometric.<sup>259,260</sup> Similarly, the orthologues of human PRMT1, PRMT3 and PRMT5 were detected in a punctate pattern along the length and on the tip of flagella in *Chlamydomonas*, a model for human ciliary disease, suggesting that PRMTs are cargo of the intraflagellar transport.<sup>261</sup> PRMT9, was not detected in these ciliary proteasome studies. However, this result does not exclude the possibility that PRMT9 is involved in the biogenesis/functioning of the cilia, since other ciliopathy genes, for example *LZTFL1*, are also not localized in the cilia.<sup>262</sup> In addition, transcriptomic analysis and subsequent RT-qPCR experiments revealed an upregulation of *GLI2* and *RPGRIP1L* in affected fibroblasts under serum starved conditions. Variations in *RPGRIP1L* are known to cause ciliopathies associated with strong brain anomalies, such as Joubert Syndrome (MIM: 611560) and Meckel-Gruber syndrome (MIM: 611561). The RPGRIP1L protein is located in the ciliary transition zone at the base of the basal body and responsible for the organized entrance and exit of proteins.<sup>263</sup> Primary cilia have been shown to play an important role in forebrain patterning *via* the regulation of the Hedgehog (Hh) pathway, in which GLI2 plays a key role.<sup>263</sup> Thus, the regulated expression of both genes, *RPGRIP1L* and *GLI2*, is essential for the function of the primary cilia and the development of the forebrain.<sup>263</sup>

Furthermore, due to the strong interaction of PRMT9 with the splice factor SAP145, PRMT9 was suggested to regulate alternative splicing. A protein complex consisting of PRMT9, SAP145 and SAP49 is most likely formed in the cytoplasm and followed by the dimethylation of SAP145 through PRMT9. SAP145 as well as SAP49 are core components of the nuclear U2 snRNP that is involved in splicing and 3' processing of pre-mRNAs.<sup>167</sup> To find targets of PRMT9 regulated splicing, Yang and colleagues compared RNA sequencing data of control knockdown, and esiRNA mediated PRMT9 knockdown HeLa cells, and identified splice variants in several ID candidate genes like *NDUFS2*, *WAC*, *BCOR* and *EEF1B2* by RNA sequencing.<sup>167,234</sup> *WAC* (*WW Domain Containing Adaptor With Coiled-Coil*), for example, is associated with similar symptoms seen in our patient: mild to severe ID, autism, behavior abnormalities, variable congenital anomalies, dysmorphic features and hypotonia.<sup>234</sup> To determine alternative splice events in the cells of our probands, we examined the transcriptome in rich (+ FCS) and serum starved (-FCS) conditions to induce ciliogenesis. Although we were unable to find the same splice variants reported by Yang *et al.* in HeLa cells, we found a number of other interesting alternative spliced genes in the skin fibroblasts of our probands, suggesting that *PRMT9* associated alternative

splicing might vary between tissues and cell types. The alternative spliced gene *SCRG1*, e.g. encodes a protein preferentially expressed in the central nervous system and associated with neurodegenerative changes in transmissible spongiform encephalopathies (TSE).<sup>246</sup> Another interesting candidate is *PDGFRB*, encoding a protein implicated in neurogenesis, synaptogenesis and modulation of ligand-gated ion channels.<sup>245</sup> However, further experiments (RT-PCRs) have to be performed to validate those alternative splice events.

In addition, the investigation of differentially expressed genes in both, serum rich (+FCS) and ciliated (-FCS) condition, gave us hints in which pathways linked to intellectual disability and autism PRMT9 might be involved. Interestingly, we observed an increase of DEGs in patient's fibroblasts cultured in serum starved conditions (-FCS) compared to serum rich conditions (+FCS), highlighting a potential role of PRMT9 during ciliogenesis. Overall, we observed a large number of DEGs associated with ID and autism, such as *SHROOM4*, *SATB2*, *RB38* or *SHANK2*<sup>20,240</sup> in both cell culture conditions. Among others, the most highly overexpressed gene is *PCDH10* that encodes a cadherin superfamily protein normally expressed in the basolateral amygdala, a brain region implicated in autistic relevant behavior<sup>264,265</sup>. The *pcdh10*<sup>-/-</sup> mice has been shown to be crucial for axon outgrowth in the forebrain.<sup>266</sup> Moreover, homozygous deletions in *PCDH10* have been identified in autistic children.<sup>267</sup>

Moreover, using the DAVID annotation tool we also observed an enrichment of DEGs in patients' cells (in both serum rich and ciliated conditions) implicated in biological processes linked to autism and ID. In +FCS conditions e.g. six DEGs (*AMBP*, *CD44*, *PCDH10*, *COL28A1*, *CDH4*, *THBS4*) associated to cell adhesion were identified. Synaptic cell adhesions molecules are involved in the cell-cell recognition during synapse formation and are strongly associated with ID and autism. Another enriched GO term was the negative regulation of transcription from RNA polymerase II promoter (DEGs: *MDFI*, *MSX2*, *HHEX*, *DLX1*, *SATB2*, *DACT1*, *HCLS1*). In the past many genes implicated in transcriptional regulation, such as transcription factors and chromatin modifiers, have been associated with ID and autism.<sup>47</sup> The transcription factor *SATB2* e.g. is strongly associated to ID and autism.<sup>268</sup> In addition to the enriched GO-terms in +FCS conditions, in ciliary conditions the enrichment of genes involved in the actin cytoskeleton organization (DEGs: *GAS2L3*, *RND3*, *SHROOM4*, *TESK2*, *SH2B2*, *ITPKA*, *FGD4*) , the regulation of synaptic plasticity (DEGs: *LZTS1*, *SYNGAP1*, *VGF*, *ITPKA*) and small GTPase mediated signal transduction (DEGs: *RND3*, *RAB32*, *RALGPS2*, *PLCE1*, *RAB39B*, *RAB38*, *DOCK11*, *DOCK4*) was observed. The cytoskeleton e.g. gives neurons its shape and stabilizes the entire cell. Moreover, the cytoskeletal organization is important during axon outgrowth and synapse formation and thus important for remodeling of synaptic connections.<sup>36</sup> Also, genes involved in the GTPase signaling pathway have been shown be involved in the organization of the actin cytoskeleton and hence have an impact on the structure and function of dendrites and synapses.<sup>14</sup> Moreover they are implicated in the intracellular vesicle trafficking in neurons, such as *RAB39B*<sup>14,40</sup>

The *MZprmt9*<sup>-/-</sup> mutant zebrafish larvae did not differ from sibling counterpart and reach adulthood normally. No brain defect was observed during the first days of development. However, this does not exclude that subtle modifications occur at later stage. More investigation would be required in order to map the different brain areas of the adult in order to detect anomalies. The variation within the *prmt9* gene leads to NMD. This was shown to frequently trigger genetic compensation. We cannot exclude that a gene takes over the role *prmt9* to minimize the effect of the deletion. RNAseq could reveal the overexpression of compensatory genes, and thus help to understand to molecular pathway implicated in this disease. This could also explain the phenotypic difference observed between *prmt9* morphants and mutants. The curved body axis observed in *prmt9* morphants might indicate a potential impairment of the primary cilia, but to further investigate this hypothesis a potential off-target effect of the morpholino oligonucleotide has to be excluded. The advantages and disadvantages of morpholinos and CRISPR mutants to study the function of a gene and the different results that we obtained using both strategies will be further discussed in the **General Discussion**.

In addition, differences between wild-type and mutant behavioral activities indicate a link between *PRMT9* and ASD. Although the time spend in the conspecific sector did not differ from wild type fish, *MZprmt9*<sup>-/-</sup> mutants showed a dramatical decrease in velocity, a behavior that was previously observed in autism zebrafish models, such as *shank3b*<sup>-/-</sup> mutants or *syngap1b* morphants.<sup>251,269</sup> In the future, it would be interesting to generate a missense variation within *prmt9* as this is described to be less prone to genetic compensation, and thus result in a more drastic phenotype.

In conclusion, we have identified novel biallelic LoF variations in the *PRMT9* gene. We definitely confirm *PRMT9* as disease gene in syndromic intellectual disability. The variations affect the function of PRMT9 in dimethylating its partner and suggest a novel role for PRMT9 in cilia biogenesis and function.

4.1.5 Supplementary Data – Project 1 (S1)

Supplementary Figures

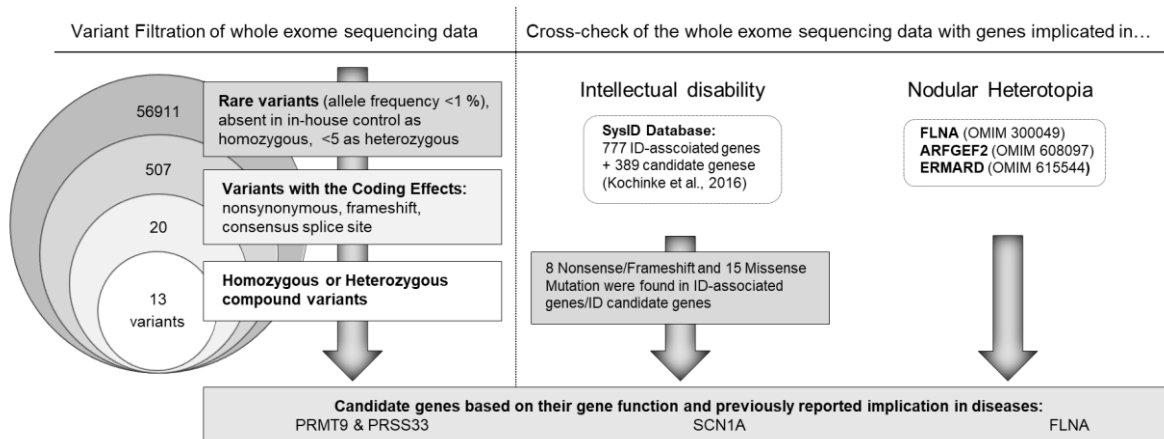


Figure S1.1 Variation filtration from whole exome sequencing data of the index patient (A.II-1) and identification of possible pathogenic variants using VaRank.

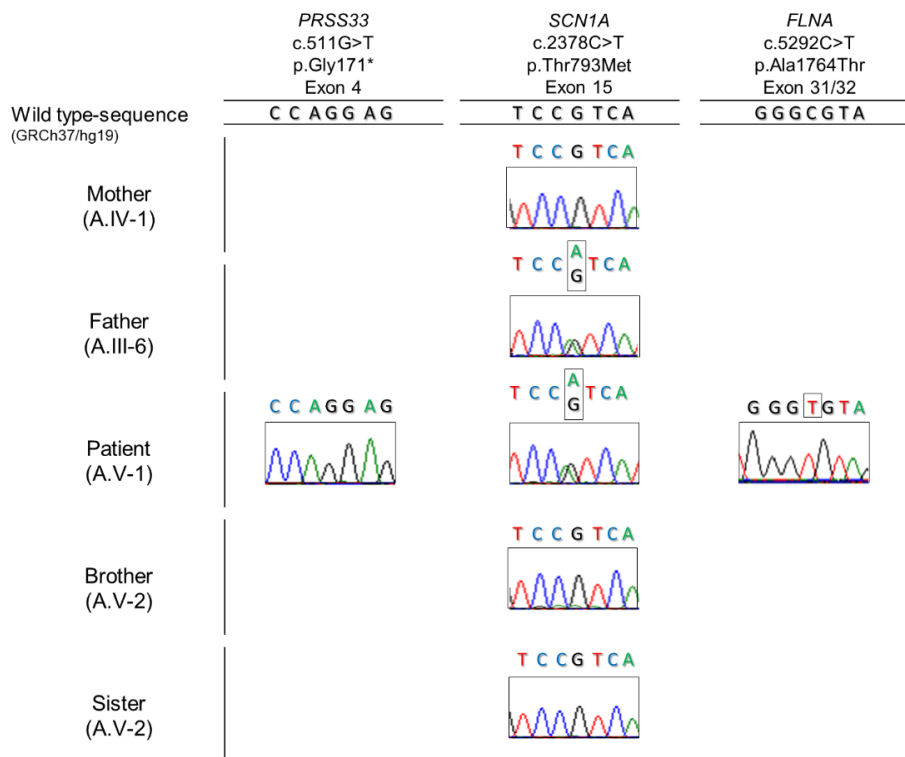
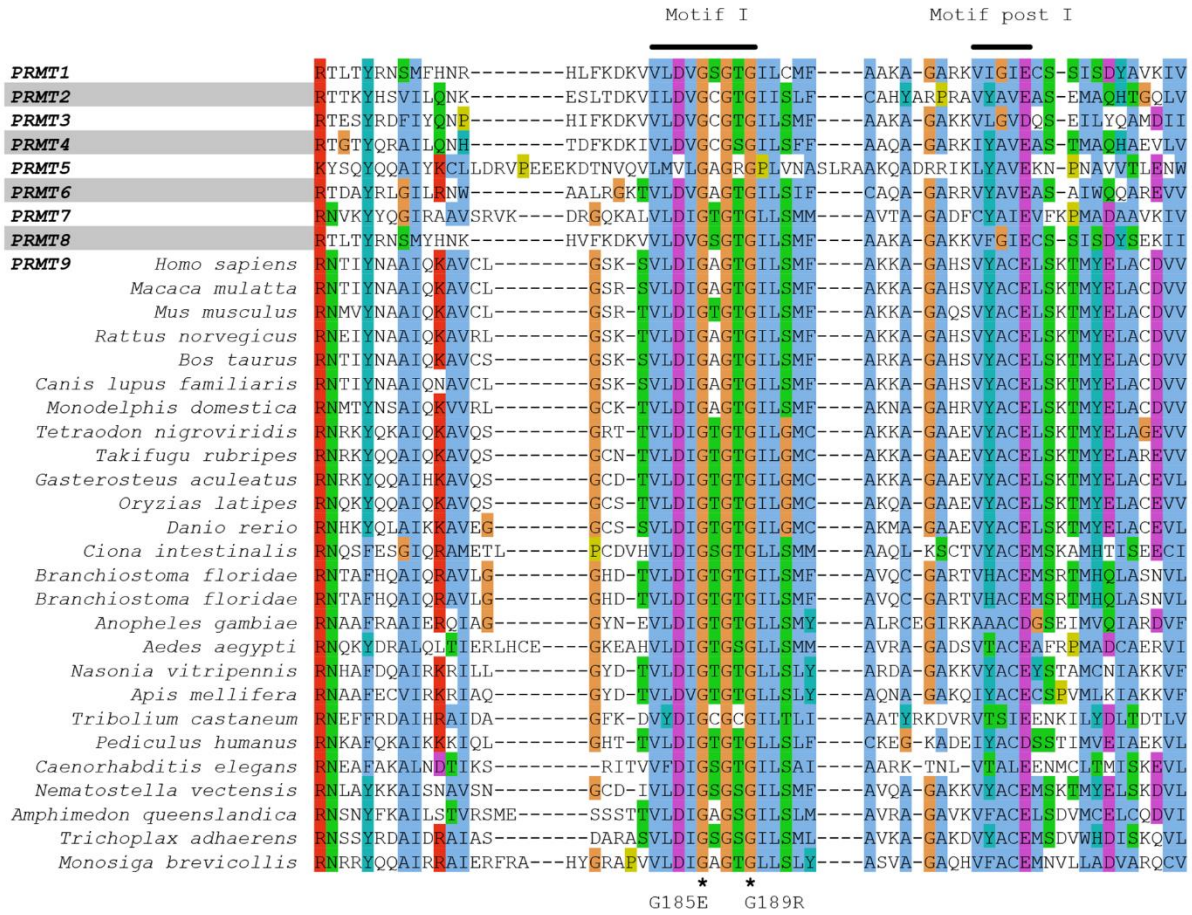


Figure S1.2 Exclusion of potential candidate genes via Sanger Sequencing.

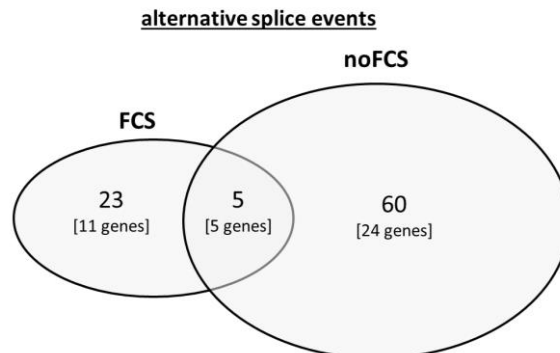
(A-C) Chromatograms of potential pathogenic gene variants in three candidate genes were generated in the patient and in family members. Nucleotides differing from the wildtype sequence (GRCh37) are framed. (A) By WES, a variation in the gene *PRSS33* was identified. Sanger Sequencing revealed that the variation was a false positive call. (B) The variation identified in the gene *SCN1A* was excluded, as it was also present in the father. (C) A single homozygous missense variant c.5292C>T in the gene *FLNA* was found in the patient by Sanger Sequencing, but was excluded because of its high allele frequency of 2 % (ExAC Browser).





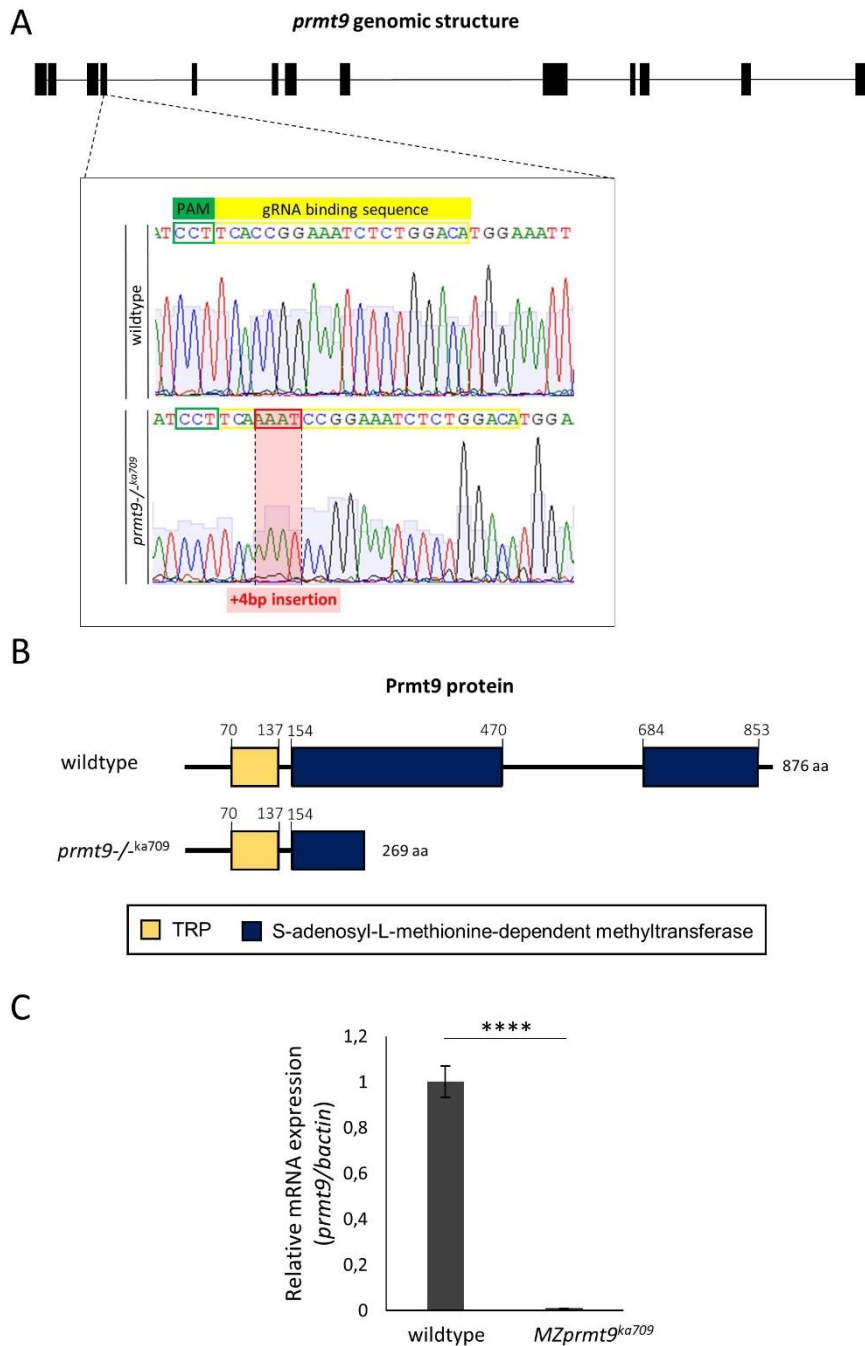
**Figure S1.3 Multiple protein sequences alignment of human PRMTs together with metazoan PRMT9.**

Proteins were extracted from the eggNOG database (v 4.5.1) <sup>1</sup> and from the UniProt database <sup>2</sup>. Sequences were aligned using Clustal O as a guide <sup>3</sup> and then manually refined. Species were ordered according to the NCBI taxonomy thanks to the phyloT online tool (<https://phylo.t.biobyte.de/>). Alignment is displayed thanks to jalview <sup>4</sup>. The 2 missenses are highlighted under the alignment with a star. Their position are within the critical Motif I of the PRMT catalytic site.



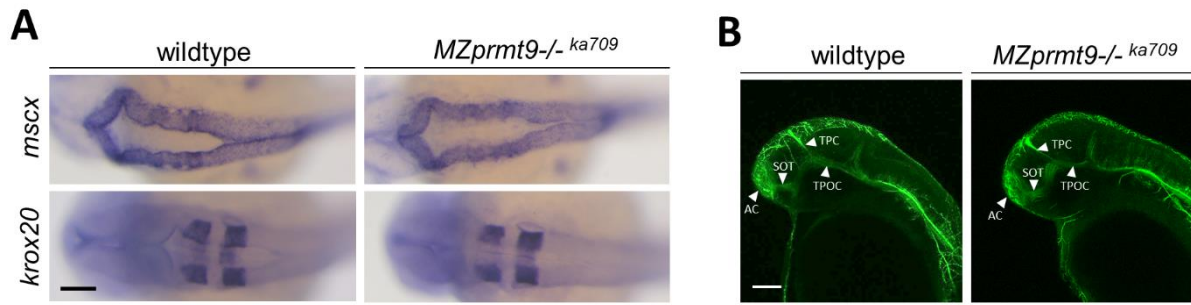
**Figure S1.4. Biallelic PRMT9 variations affect the alternative splicing of genes.**

Prior to RNA extraction and sequencing, control cells and patients fibroblasts (individual AII-1, CII-15 and CII-2) were cultured in normal conditions (+FCS) and ciliated conditions (-FCS). To induce primary cilium formation, the cells were deprived of serum by growth for 48h in DMEM with 1 % Pen/Strep. Using Leafcutter (filter: p.adj.<0.01; log<sub>2</sub>FC >0.6 or <-0.6; ΔΨ<-0.5 or ΔΨ>0.5) and JunctionSeq (filter: log<sub>2</sub>FC >3 or <-3; p.adj. <0.01; dispersion < 0.05), we observed 16 alternative spliced genes (13 detected by JunctionSeq and 3 by Leafcutter) in patients' cells cultivated with FCS and 29 alternative spliced genes (13 detected by JunctionSeq and 3 by Leafcutter) in patients' cells cultivated without FCS.



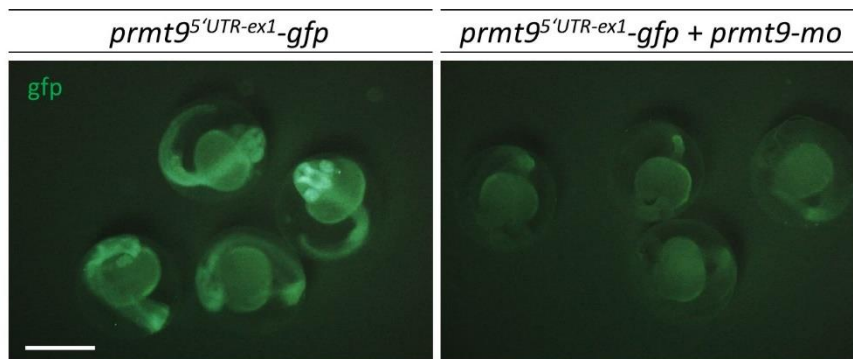
**Figure S1.5 Knockout of *prmt9* in zebrafish using the CRISPR/Cas9 system.**

**(A)** The guide RNA binds a sequence within exon 4 of the *prmt9* gene and led to an insertion of four base pairs resulting in a frameshift and a premature stop codon. **(B)** Schematic representation of the Prmt9 protein structure of wildtype and *prmt9*<sup>-/-</sup> mutants. **(C)** Real-time qPCR analyses of the expression of *PRMT9*, normalized to  $\beta$ -actin in wildtype and *prmt9*<sup>-/-</sup> mutants. Error bars represent standard deviation of biological triplicates. n= 50 embryos (3 dpf)/replicate. Statistical significance was determined using the unpaired, t-Test, ns = non-significant. \* p < 0.05. \*\* p < 0.01. \*\*\* p < 0.001. \*\*\*\* p < 0.0001.

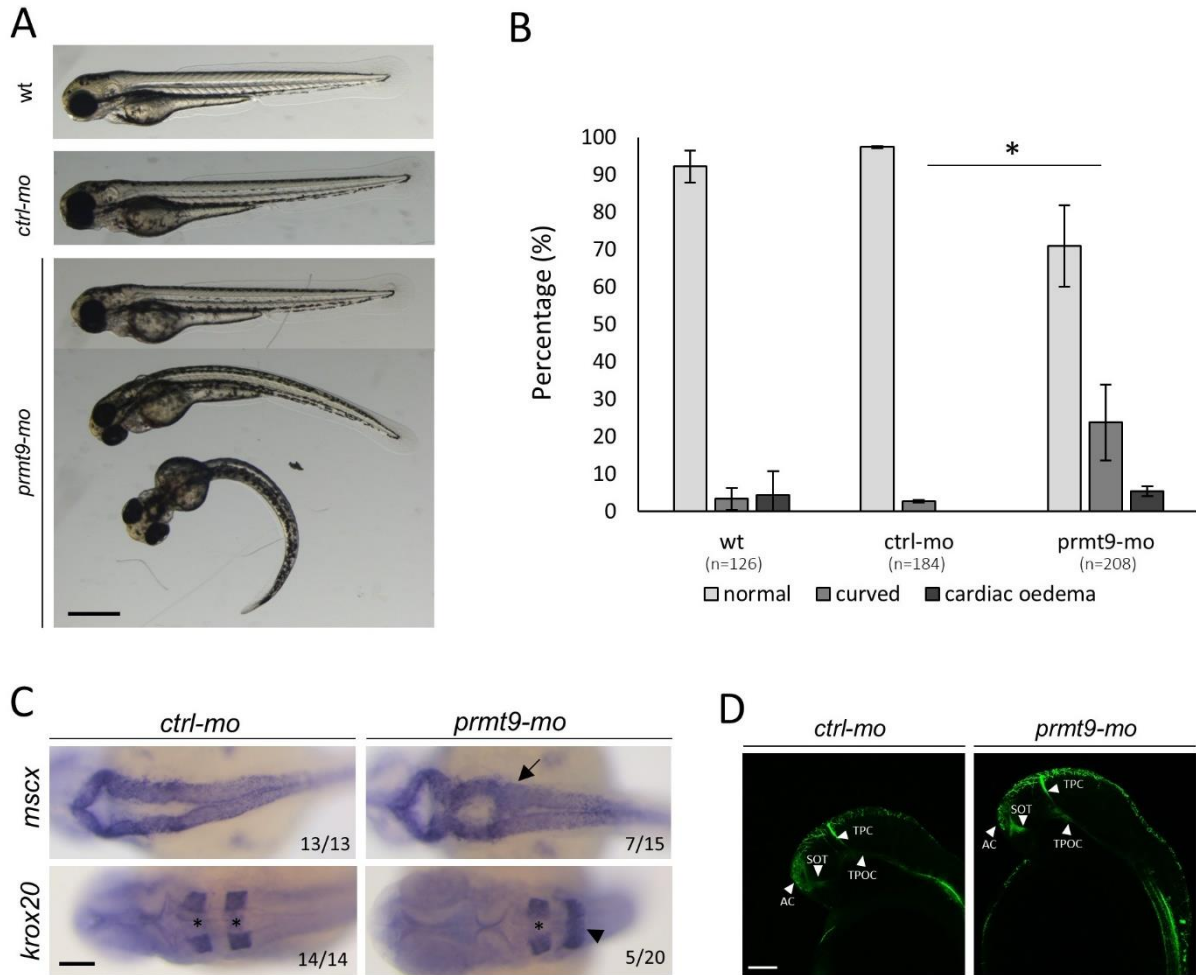


**Figure S1.6 No differences during brain development were observed in *MZprmt9*<sup>-/-</sup> zebrafish embryos.**

**(A)** Whole-mount *in situ* hybridizations at 24 hpf. Dorsal view of embryos stained with probes against the brain marker genes *mscx* (hindbrain, rhombencephalon) and *krox20* (midbrain-hindbrain boundary). Scale bar 100 μm. n=15. **(B)** Axon tracts of wildtype and mutant fish were immunostained with an antibody against acetylated tubulin (35 hpf). AC anterior commissures, SOT supra optic tract, TPC tract of the posterior commissure, TPOC tract of the post optic commissure. Scale bar = 100 μm.



**Figure S1.7 Morpholino efficiency of *prmt9-mo* was assessed by co-injecting a *prmt9* N-terminal sequence fused to GFP. Scale bar = 500 μm.**



**Figure S1.8** *prmt9-mo* injected embryos present a curved body axis (3dpf) and an impaired brain morphogenesis (24 hpf).

**(A)** Representative images of 3 dpf old wildtype, *ctrl-mo* and *prmt9-mo* injected zebrafish embryos. Scale bar 500  $\mu\text{m}$ .

**(B)** Percentage of embryos with a curved body axis or cardiac oedema. Compared to control injected fish significant more embryos injected with *prmt9-mo* present a curved body axis.

**(C)** Whole-mount *in situ* hybridizations at 24 hpf. Dorsal view of embryos stained with probes against the brain marker genes *mscx* (hindbrain, rhombencephalon) and *krox20* (midbrain-hindbrain boundary).

While the brain ventricle is open in all control injected embryos (*ctrl-mo*, *mscx* n=13, *krox20* n=14), the brain ventricle is still partly closed in 25 % (*krox20* n=20, arrowhead) to 47 % (*mscx* n=15, arrow) of *prmt9-mo* injected embryos. Scale bar 100  $\mu\text{m}$ .

**(D)** Axon tracts of control (*ctrl-mo*) and *prmt9-mo* injected embryos were immunostained with an antibody against acetylated tubulin (35 hpf). AC anterior commissures, SOT supra optic tract, TPC tract of the posterior commissure, TPOC tract of the post optic commissure. Scale bar = 100 $\mu\text{m}$ .

## Supplementary Tables

**Table S1.1 Homozygous regions specific to the affected individual A.II-1.** Homozygosity mapping has been performed on one affected individual using the Affymetrix 50K SNP array.

Chromosome	Zone	250K Affymetrix Array	Genomic position (hg19)	Candidate Gene
Chr 1	1	rs10493519- rs413853	74,178107Mb to 75,291698Mb	
	2	rs10493622 - rs10493662	79,484397Mb to 80,809034Mb	
Chr 2	1	rs10496466 -rs332920	115,33725Mb to 117,881707Mb	
	2	rs3795865- rs1878199	187,500682Mb to 189,831353Mb	
	3	rs10497909-rs1074158	209,539426Mb to 211,175043Mb	
Chr 4	1	rs6833118 - rs10488949	6,835067Mb to 10,589824Mb	
	2	rs6841592 - rs2412628	38,964365Mb to 55,225917Mb	
	3	rs2597883 - rs9312352	111,688738Mb to 168,706813Mb	<i>PRMT9</i>
	4	rs1426935 - rs10520333	174,609314Mb to 176,129327Mb	
Chr 5	1	rs331670 - rs9312746	8,108398Mb to 10,916699Mb	
Chr 6	1	rs1857859 - rs596621	100,894587Mb to 102,055493Mb	
	2	rs699418 - rs6906869	128,255410Mb to 129,727172Mb	
Chr 8	1	rs10504631 - rs1607547	77,538487Mb to 78,658127Mb	
Chr 9	1	rs10491515 - rs10513456	125,387853Mb to 128,719849Mb	
Chr 10	1	rs9299542 - rs1913760	79,941665Mb to 82,344676Mb	
Chr 11	1	rs10501217 - rs4756656	40,092369Mb to 41,586140Mb	
Chr 13	1	rs1360367 - rs9318616	78,682863Mb to 79,801863Mb	
	2	rs4128156 - rs1602222	84,706661Mb to 85,768963Mb	
Chr 14	1	rs10483780 - rs181570	65,980170Mb to 68,275342Mb	
Chr 15	1	rs744319 - rs2017566	57,881700Mb to 58,505040Mb	
	2	rs3844975 - rs718457	87,720711Mb to 89,785811Mb	
Chr 17	1	chromosome start - rs4239056	1bp to 5,324406Mb	
Chr 18	1	rs1443454 - rs10502520	25,554315Mb to 26,801474Mb	
Chr 19	1	rs8109402 - rs8106841	42,131442Mb to 46,327709Mb	
Chr 20	1	rs202518 - rs910952	1,705474Mb to 4,193694Mb	

**Table S1.2 Differentially expressed genes (+FCS)**

Ensembl gene ID	Gene name	Description	log2 FC	Adj p-value
<b>Upregulated with log2FC&gt;1 and adjusted p-value &lt;0,05</b>				
ENSG00000042980	PCDH10	protocadherin 10	7,4	2,03E-05
ENSG00000104826	SYT15	synaptotagmin XV	6,1	5,57E-03
ENSG00000004777	CD200	CD200 molecule	7,4	3,95E-05
ENSG00000090612	A2M	alpha-2-macroglobulin	6,4	5,89E-04
ENSG00000057019	LHX8	LIM homeobox 8	5,9	1,32E-02
ENSG00000088538	COL11A1	collagen, type XI, alpha 1	5,4	3,28E-02
ENSG00000064300	TM6SF1	transmembrane 6 superfamily member 1	5,6	3,92E-04
ENSG00000113296	HCLS1	hematopoietic cell-specific Lyn substrate 1	4,3	1,48E-02
ENSG00000112246	AGT	angiotensinogen (serpin peptidase inhibitor, clade A, member 8)	4,1	3,18E-02
ENSG00000109861	PLXDC2	plexin domain containing 2	4,1	1,02E-02
ENSG00000111817	PPAPDC1A	phosphatidic acid phosphatase type 2 domain containing 1A	3,9	8,93E-06

Supplementary Data – Project 1

Table S1.2 (continued)

Ensembl gene ID	Gene name	Description	log2 FC	Adj p-value
<b>Upregulated with log2FC&gt;1 and adjusted p-value &lt;0,05</b>				
ENSG00000101115	DACT1	dishevelled-binding antagonist of beta-catenin 1	4,3	1,99E-04
ENSG00000052850	MDFI	MyoD family inhibitor	5,0	8,06E-04
ENSG00000117289	RP11-13P5.2		3,5	3,30E-02
ENSG00000112559	RTN1	reticulon 1	3,7	1,98E-04
ENSG00000120594	FNDC1	fibronectin type III domain containing 1	3,4	2,05E-05
ENSG00000103254	TSPAN2	tetraspanin 2	4,0	2,40E-03
ENSG00000064886	C11orf96	chromosome 11 open reading frame 96	4,6	2,26E-02
ENSG00000120645	CDH4	cadherin 4, type 1, R-cadherin (retinal)	3,2	1,02E-02
ENSG00000132004	SGK223	Tyrosine-protein kinase Sgk223	2,7	1,49E-02
ENSG00000104899	CASC15	cancer susceptibility candidate 15 (non-protein coding)	3,6	1,55E-02
ENSG00000159176	DSP	desmoplakin	2,2	1,31E-02
ENSG00000116141	CTB-78F1.2		3,0	2,52E-02
ENSG00000129682	SHROOM4	shroom family member 4	2,4	1,33E-04
ENSG00000144749	MYO1E	myosin IE	1,7	1,07E-02
ENSG00000154330	TPST2	tyrosylprotein sulfotransferase 2	1,6	3,06E-03
<b>Downregulated with log2FC&lt;-1 and adjusted p-value &lt;0,05</b>				
ENSG00000166426	SATB2	SATB homeobox 2	-1,0	1,55E-02
ENSG00000168646	TIPARP	TCDD-inducible poly(ADP-ribose) polymerase	-1,1	4,59E-03
ENSG00000168779	DCBLD2	discoidin, CUB and LCCL domain containing 2	-1,3	1,56E-02
ENSG00000172548	CD44	CD44 molecule (Indian blood group)	-1,2	1,85E-02
ENSG00000171757	ADSSL1	adenylosuccinate synthase like 1	-1,3	3,13E-02
ENSG00000264575	DLX1	distal-less homeobox 1	-3,5	5,62E-03
ENSG00000172247	PSD3	pleckstrin and Sec7 domain containing 3	-1,5	7,80E-03
ENSG00000174564	AHR	aryl hydrocarbon receptor	-1,9	4,11E-03
ENSG00000172789	ZNF521	zinc finger protein 521	-1,7	5,15E-04
ENSG00000185133	SIM1	single-minded family bHLH transcription factor 1	-2,2	9,74E-03
ENSG00000175782	MSX2	msh homeobox 2	-2,1	2,10E-02
ENSG00000186832	FAM218A	family with sequence similarity 218, member A	-2,4	2,13E-03
ENSG00000182580	HHEX	hematopoietically expressed homeobox	-2,3	1,57E-02
ENSG00000187398	COLEC12	collectin sub-family member 12	-2,6	1,11E-03
ENSG00000184922	AIM1	absent in melanoma 1	-2,8	5,07E-03
ENSG00000215018	CD7	CD7 molecule	-3,5	1,34E-02
ENSG00000233368	APOD	apolipoprotein D	-3,9	5,27E-04
ENSG00000186530	DENND2A	DENN/MADD domain containing 2A	-3,1	2,09E-02
ENSG00000198056	COL28A1	collagen, type XXVIII, alpha 1	-3,4	2,08E-02
ENSG00000187231	TMEM26	transmembrane protein 26	-3,2	1,62E-05
ENSG00000196924	CTSC	cathepsin C	-3,5	2,11E-02
ENSG00000218052	ADAM28	ADAM metalloproteinase domain 28	-3,9	2,29E-02
ENSG00000185100	RAB38	RAB38, member RAS oncogene family	-3,3	1,63E-02
ENSG00000183171	ADRB2	adrenoceptor beta 2, surface	-3,3	1,33E-02
ENSG00000258537	AMBP	alpha-1-microglobulin/bikunin precursor	-4,7	1,73E-02
ENSG00000240476	NPTX1	neuronal pentraxin I	-4,4	3,90E-02
ENSG00000266037	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)	-5,0	7,64E-03
ENSG00000205426	WNT16	wingless-type MMTV integration site family, member 16	-4,0	1,34E-02
ENSG00000257526	THBS4	thrombospondin 4	-5,0	2,22E-03
ENSG00000221319	FGF13	fibroblast growth factor 13	-4,4	1,40E-02
ENSG00000271361	hsa-mir-6723	hsa-mir-6723	-5,7	4,35E-03
ENSG00000231852	CLEC3B	C-type lectin domain family 3, member B	-5,1	1,10E-02
ENSG00000272343	TAS1R1	taste receptor, type 1, member 1	-6,4	4,27E-04
ENSG00000264462	ADH1B	alcohol dehydrogenase 1B (class I), beta polypeptide	-6,1	1,79E-06
ENSG00000179772	HOXA13	homeobox A13	-5,1	6,12E-03
ENSG00000228502	VIT	vitrin	-6,2	2,02E-02



**Table S1.3 Functional Annotation (+FCS).** Data were generated using the online tool DAVID.

GO term: Biological Process	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0050731 ~positive regulation of peptidyl-tyrosine phosphorylation	5	8,47E+14	9,47E+10	CD44, AGT, HCLS1, HGF, THBS4	50	82	16792	2,05E+16	0,04	0,04	1,4E-01
GO:0014068 ~positive regulation of phosphatidylinositol 3-kinase signaling	4	6,78E+15	8,98E+10	WNT16, AGT, HCLS1, HGF	50	65	16792	2,07E+16	0,35	0,19	1,28E+16
GO:0010951 ~negative regulation of endopeptidase activity	4	6,78E+15	5,28E-03	AMBP, A2M, AGT, COL28A1	50	121	16792	1,11E+16	0,92	0,57	7,31E+15
GO:0048704 ~embryonic skeletal system morphogenesis	3	5,08E+15	5,77E-03	MDFI, SATB2, COL11A1	50	39	16792	2,58E+16	0,94	0,50	7,96E+15
GO:0007155 ~cell adhesion	6	1,02E+15	1,06E-02	AMBP, CD44, PCDH10, COL28A1, CDH4, THBS4	50	459	16792	4,39E+15	0,99	0,64	1,42E+16
GO:0001570 ~vasculogenesis	3	5,08E+15	1,16E-02	HHEX, MYO1E, TIPARP	50	56	16792	1,80E+15	1,00	0,61	1,54E+16
GO:0048619 ~embryonic hindgut morphogenesis	2	3,39E+15	1,16E-02	DACT1, HOXA13	50	4	16792	1,68E+16	1,00	0,55	1,54E+16
GO:0016525 ~negative regulation of angiogenesis	3	5,08E+15	1,41E-02	HHEX, AGT, THBS4	50	62	16792	1,63E+15	1,00	0,58	1,84E+16
GO:0000122 ~negative regulation of transcription from RNA polymerase II promoter	7	1,19E+16	1,77E-02	MDFI, MSX2, HHEX, DLX1, SATB2, DACT1, HCLS1	50	720	16792	3,27E+16	1,00	0,62	2,26E+16
GO:0002576 ~platelet degranulation	3	5,08E+15	3,64E-02	A2M, CLEC3B, HGF	50	103	16792	9,78E+15	1,00	0,83	4,12E+15
GO:0006807 ~nitrogen compound metabolic process	2	3,39E+15	4,01E-02	MYO1E, TIPARP	50	14	16792	4,80E+15	1,00	0,83	4,44E+15
GO:0002063 ~chondrocyte development	2	3,39E+15	4,29E-02	MSX2, COL11A1	50	15	16792	4,48E+16	1,00	0,83	4,66E+15
GO:0008585 ~female gonad development	2	3,39E+15	4,29E-02	TIPARP, LHX8	50	15	16792	4,48E+16	1,00	0,83	4,66E+15
GO:0071310 ~cellular response to organic substance	2	3,39E+15	4,85E-02	SATB2, CLEC3B	50	17	16792	3,95E+16	1	0,84	5,09E+16

Table S1.3 (continued)

GO term: Biological Process	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0002076 ~osteoblast development	2	3,39E+15	4,85E-02	MSX2, SATB2	50	17	16792	3,95E+16	1	0,84	5,09E+16
GO:0046329 ~negative regulation of JNK cascade	2	3,39E+15	6,23E-02	AMBP, DACT1	50	22	16792	3,05E+16	1	0,89	6,02E+15
GO:0048863 ~stem cell differentiation	2	3,39E+15	7,59E-02	MSX2, A2M	50	27	16792	2,49E+16	1	0,92	6,77E+14
GO:0007156 ~homophilic cell adhesion via plasma membrane adhesion molecules	3	5,08E+15	7,77E-02	PCDH10, CD200, CDH4	50	158	16792	6,38E+15	1	0,91	6,86E+15
GO:0048008 ~platelet-derived growth factor receptor signaling pathway	2	3,39E+15	8,13E-02	MYO1E, TIPARP	50	29	16792	2,32E+16	1	0,91	7,03E+14
GO:0007568 ~aging	3	5,08E+15	8,36E-02	APOD, AGT, CTSC	50	165	16792	6,11E+15	1	0,90	7,14E+15
GO:0030177 ~positive regulation of Wnt signaling pathway	2	3,39E+15	1,00E-01	HHEX, DACT1	50	36	16792	1,87E+16	1	0,93	7,79E+14



Supplementary Data – Project 1

**Table S1.4 Differentially expressed genes (-FCS/ ciliary conditions)**

Ensembl gene ID	Gene name	Description	log2 FC	Adj p-value
<b>Upregulated with log2FC&gt;1 and adjusted p-value &lt;0,05</b>				
ENSG0000004799	SHANK2	SH3 and multiple ankyrin repeat domains 2	5,4	1,9E-03
ENSG00000005102	CD200	CD200 molecule	6,4	3,1E-04
ENSG00000019582	PRR15	proline rich 15	5,5	1,4E-02
ENSG00000019991	SYNDIG1	synapse differentiation inducing 1	5,7	6,6E-03
ENSG00000024526	PSTPIP1	proline-serine-threonine phosphatase interacting protein 1	5,9	7,7E-05
ENSG00000028137	SULT1A1	sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1	6,3	1,5E-04
ENSG00000042062	PTPRD	protein tyrosine phosphatase, receptor type, D	4,7	2,0E-02
ENSG00000046889	MYH1	myosin, heavy chain 1, skeletal muscle, adult	6,0	1,1E-02
ENSG00000047648	SLC16A9	solute carrier family 16, member 9	4,9	1,1E-03
ENSG00000051341	HCLS1	hematopoietic cell-specific Lyn substrate 1	5,0	2,8E-04
ENSG00000062582	A2M	alpha-2-macroglobulin	4,3	2,5E-02
ENSG00000065320	DEPTOR	DEP domain containing MTOR-interacting protein	3,9	1,5E-02
ENSG00000067177	ISLR2	immunoglobulin superfamily containing leucine-rich repeat 2	4,9	2,3E-03
ENSG00000068784	PCDH10	protocadherin 10	4,4	1,1E-02
ENSG00000069188	RP11-839D17.3		4,8	3,2E-02
ENSG00000069702	CTB-78F1.2		4,1	4,8E-04
ENSG00000070159	RP11-13P5.2		4,1	4,8E-03
ENSG00000070614	ANO1	anoctamin 1, calcium activated chloride channel	4,4	3,2E-02
ENSG00000073060	RP11-536K7.3		3,3	1,5E-02
ENSG00000073464	AC093850.2		4,1	2,8E-02
ENSG00000074527	TSPAN2	tetraspanin 2	4,7	9,7E-05
ENSG00000075651	FNDC1	fibronectin type III domain containing 1	3,8	2,4E-07
ENSG00000076641	MDFI	MyoD family inhibitor	4,9	3,8E-04
ENSG00000078295	RTN1	reticulon 1	3,5	5,3E-04
ENSG00000079459	SEL1L3	sel-1 suppressor of lin-12-like 3 (C. elegans)	3,4	1,9E-02
ENSG00000081277	KLHL2P1	kelch-like family member 2 pseudogene 1	4,3	2,5E-02
ENSG00000081803	RP11-261P9.4		4,0	6,3E-03
ENSG00000085377	PCDH7	protocadherin 7	6,1	1,2E-02
ENSG00000086289	AC007879.7		3,5	5,8E-05
ENSG00000087111	TENM4	teneurin transmembrane protein 4	3,5	4,3E-03
ENSG00000089685	SORBS1	sorbin and SH3 domain containing 1	4,3	1,6E-02
ENSG00000090487	POSTN	periostin, osteoblast specific factor	3,8	2,7E-02
ENSG00000090889	AC002480.3		3,9	4,8E-03
ENSG00000091844	DACT1	dishevelled-binding antagonist of beta-catenin 1	3,1	1,6E-02
ENSG00000095637	PLCE1	phospholipase C, epsilon 1	3,4	3,8E-02
ENSG00000096696	DPF1	D4, zinc and double PHD fingers family 1	4,1	4,6E-04
ENSG00000099139	CDH4	cadherin 4, type 1, R-cadherin (retinal)	3,1	1,4E-02
ENSG00000100167	TMEM92	transmembrane protein 92	3,5	2,2E-02
ENSG00000100784	LZTS1	leucine zipper, putative tumor suppressor 1	3,1	6,4E-03
ENSG00000101017	C8orf47	chromosome 8 open reading frame 47	3,3	3,0E-04
ENSG00000101096	SGK223	Tyrosine-protein kinase Sgk223	2,9	2,9E-03
ENSG00000101282	RAB39B	RAB39B, member RAS oncogene family	3,3	5,2E-03
ENSG00000101447	ID4	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein	2,8	1,6E-02

Supplementary Data – Project 1

Table S1.4 (continued)

Ensembl gene ID	Gene name	Description	log2 FC	Adj p-value
<b>Upregulated with log2FC&gt;1 and adjusted p-value &lt;0,05</b>				
ENSG00000102804	KIAA1549L	KIAA1549-like	2,8	1,0E-02
ENSG00000103150	SERPINB7	serpin peptidase inhibitor, clade B (ovalbumin), member 7	2,9	4,2E-02
ENSG00000103569	MYOM2	myomesin 2	2,9	4,6E-04
ENSG00000104147	SEMA7A	semaphorin 7A, GPI membrane anchor (John Milton Hagen blood group)	2,5	2,3E-04
ENSG00000104524	BCAT1	branched chain amino-acid transaminase 1, cytosolic	2,6	1,9E-02
ENSG00000104951	GDAP1	ganglioside induced differentiation associated protein 1	3,7	1,4E-03
ENSG00000105499	GLI2	GLI family zinc finger 2	2,6	1,1E-02
ENSG00000105516	GUSBP5	glucuronidase, beta pseudogene 5	2,7	1,8E-03
ENSG00000105711	PODNL1	podocan-like 1	2,3	3,1E-03
ENSG00000105825	PYCR1	pyrroline-5-carboxylate reductase 1	2,5	1,7E-02
ENSG00000106031	MLLT11	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 11	2,5	2,6E-03
ENSG00000106565	ADAMTS12	ADAM metalloproteinase with thrombospondin type 1 motif, 12	2,2	1,4E-03
ENSG00000106829	GPCPD1	glycerophosphocholine phosphodiesterase GDE1 homolog (S. cerevisiae)	2,0	2,5E-06
ENSG00000107317	TSC22D3	TSC22 domain family, member 3	2,2	5,1E-03
ENSG00000108231	COL1A1	collagen, type I, alpha 1	2,2	9,9E-05
ENSG00000108556	PRPS1	phosphoribosyl pyrophosphate synthetase 1	2,1	8,9E-03
ENSG00000108679	SHROOM4	shroom family member 4	2,2	4,8E-04
ENSG00000108932	RP1-191J18.66		2,2	1,1E-03
ENSG00000109061	RP5-968P14.2		2,8	2,3E-02
ENSG00000110347	SLC1A4	solute carrier family 1 (glutamate/neutral amino acid transporter), member 4	2,3	6,9E-03
ENSG00000110876	MYH10	myosin, heavy chain 10, non-muscle	2,0	1,1E-02
ENSG00000111087	MSMO1	methylsterol monooxygenase 1	2,3	2,5E-02
ENSG00000111913	ADCY7	adenylate cyclase 7	1,3	1,9E-02
ENSG00000112149	ANXA2	annexin A2	1,5	2,5E-02
ENSG00000112299	C7orf50	chromosome 7 open reading frame 50	1,3	1,9E-03
ENSG00000112984	NLN	neurolysin (metallopeptidase M3 family)	1,4	2,4E-02
ENSG00000113722	PPIC	peptidylprolyl isomerase C (cyclophilin C)	1,2	1,5E-02
ENSG00000114850	MAGOHB	mago-nashi homolog B (Drosophila)	1,4	2,3E-02
ENSG00000115163	PYROXD1	pyridine nucleotide-disulphide oxidoreductase domain 1	1,2	1,4E-02
ENSG00000115233	FLNA	filamin A, alpha	1,5	1,8E-02
ENSG00000115594	HCG18	HLA complex group 18 (non-protein coding)	1,1	1,7E-02
ENSG00000116157	RPGRIP1L	RPGRIP1-like	1,3	2,9E-03
<b>Downregulated with log2FC&gt;-1 and adjusted p-value &lt;0,05</b>				
ENSG00000117399	GLI3	GLI family zinc finger 3	-1,3	2,3E-02
ENSG00000117650	CD44	CD44 molecule (Indian blood group)	-1,4	1,0E-02
ENSG00000118193	DBP	D site of albumin promoter (albumin D-box) binding protein	-1,2	2,6E-02
ENSG00000118523	RICTOR	RPTOR independent companion of MTOR, complex 2	-1,2	1,7E-02
ENSG00000119599	NBPF10	neuroblastoma breakpoint family, member 10	-1,2	9,8E-03
ENSG00000119630	TKT	transketolase	-1,4	2,2E-03
ENSG00000119686	ARMCX1	armadillo repeat containing, X-linked 1	-1,3	1,8E-02
ENSG00000119917	AHRR	aryl-hydrocarbon receptor repressor	-1,3	4,2E-03
ENSG00000120594	SH2B2	SH2B adaptor protein 2	-1,7	2,2E-03

Supplementary Data – Project 1

Table S1.4 (continued)

Ensembl gene ID	Gene name	Description	log2 FC	Adj p-value
Downregulated with log2FC>-1 and adjusted p-value <0,05				
ENSG00000120875	ZNF385A	zinc finger protein 385A	-1,3	7,8E-03
ENSG00000121060	KLF11	Kruppel-like factor 11	-1,3	4,6E-02
ENSG00000122378	SHOX2	short stature homeobox 2	-1,9	7,7E-03
ENSG00000123485	MAN1A1	mannosidase, alpha, class 1A, member 1	-1,5	2,2E-03
ENSG00000123496	C14orf132	chromosome 14 open reading frame 132	-1,5	1,0E-04
ENSG00000123689	ACPL2	acid phosphatase-like 2	-1,6	9,5E-04
ENSG00000123975	PGPEP1	pyroglutamyl-peptidase I	-1,5	4,8E-02
ENSG00000124508	RAB32	RAB32, member RAS oncogene family	-1,5	2,9E-02
ENSG00000125246	KIAA1147	KIAA1147	-1,4	9,9E-03
ENSG00000125398	KIF13B	kinesin family member 13B	-1,4	1,9E-02
ENSG00000125730	DHRS13	dehydrogenase/reductase (SDR family) member 13	-1,7	1,8E-02
ENSG00000125869	DOCK11	dedicator of cytokinesis 11	-1,7	1,2E-02
ENSG00000128266	SHBG	sex hormone-binding globulin	-2,1	1,0E-02
ENSG00000128918	RALGPS2	Ral GEF with PH domain and SH3 binding motif 2	-1,5	4,5E-03
ENSG00000129946	TWIST2	twist family bHLH transcription factor 2	-1,9	4,4E-03
ENSG00000130052	C4orf36	chromosome 4 open reading frame 36	-1,5	7,1E-03
ENSG00000130147	GYPC	glycophorin C (Gerbich blood group)	-1,8	1,6E-02
ENSG00000130224	PELI2	pellino E3 ubiquitin protein ligase family member 2	-1,7	3,5E-02
ENSG00000131797	FAXDC2	fatty acid hydroxylase domain containing 2	-1,8	6,2E-05
ENSG00000132334	NBL1	neuroblastoma 1, DAN family BMP antagonist	-1,7	2,5E-03
ENSG00000132510	SYNGAP1	synaptic Ras GTPase activating protein 1	-2,0	1,1E-02
ENSG00000133134	H1FX	H1 histone family, member X	-1,8	2,7E-02
ENSG00000134215	ACYP1	acylphosphatase 1, erythrocyte (common) type	-2,0	1,9E-02
ENSG00000134602	GJA1	gap junction protein, alpha 1, 43kDa	-1,8	3,2E-02
ENSG00000134769	LIFR	leukemia inhibitory factor receptor alpha	-1,8	1,2E-02
ENSG00000135378	STOM	stomatin	-1,8	2,6E-02
ENSG00000135451	CNNM2	cyclin M2	-1,9	3,6E-02
ENSG00000135605	TMEM86A	transmembrane protein 86A	-1,9	7,9E-04
ENSG00000136040	TNFRSF14	tumor necrosis factor receptor superfamily, member 14	-2,1	8,8E-04
ENSG00000137094	TESK2	testis-specific kinase 2	-1,7	1,8E-02
ENSG00000137462	BCL2L11	BCL2-like 11 (apoptosis facilitator)	-1,9	3,3E-02
ENSG00000137752	RP11-804H8.6		-2,8	4,8E-03
ENSG00000138180	MSX2	msh homeobox 2	-2,2	8,3E-03
ENSG00000138646	SOCS1	suppressor of cytokine signaling 1	-2,1	2,6E-02
ENSG00000138650	DOCK4	dedicator of cytokinesis 4	-1,8	2,2E-02
ENSG00000138944	ARID5A	AT rich interactive domain 5A (MRF1-like)	-1,9	1,9E-02
ENSG00000139266	RPS6KA5	ribosomal protein S6 kinase, 90kDa, polypeptide 5	-2,2	2,4E-02
ENSG00000139428	RND3	Rho family GTPase 3	-2,2	2,1E-02
ENSG00000139910	ANGPTL2	angiopoietin-like 2	-2,0	2,5E-02
ENSG00000141198	SPATA13	spermatogenesis associated 13	-1,7	1,7E-02
ENSG00000141429	MRPS6	mitochondrial ribosomal protein S6	-2,1	2,3E-03
ENSG00000143195	IL1R1	interleukin 1 receptor, type I	-2,6	1,7E-02
ENSG00000143631	HPCAL1	hippocalcin-like 1	-2,1	9,2E-03
ENSG00000143797	MAFB	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog B	-2,0	3,2E-02

Supplementary Data – Project 1

Table S1.4 (continued)

Ensembl gene ID	Gene name	Description	log2 FC	Adj p-value
Downregulated with log2FC>-1 and adjusted p-value <0,05				
ENSG00000143850	H1FO	H1 histone family, member 0	-2,2	3,0E-02
ENSG00000144481	ETS2	v-ets avian erythroblastosis virus E26 oncogene homolog 2	-2,2	2,4E-02
ENSG00000145358	TGFA	transforming growth factor, alpha	-2,0	2,1E-02
ENSG00000145861	ANKRD9	ankyrin repeat domain 9	-1,9	1,8E-02
ENSG00000146072	AHR	aryl hydrocarbon receptor	-1,9	1,7E-02
ENSG00000146233	RP11-554A11.4		-2,4	4,9E-05
ENSG00000146374	FAM107B	family with sequence similarity 107, member B	-1,8	2,4E-02
ENSG00000146477	C1orf132	chromosome 1 open reading frame 132	-2,8	9,6E-03
ENSG00000146966	COLEC12	collectin sub-family member 12	-2,4	2,1E-03
ENSG00000150048	ZNF702P	zinc finger protein 702, pseudogene	-2,3	1,6E-02
ENSG00000151229	CREBRF	CREB3 regulatory factor	-2,6	1,9E-02
ENSG00000152092	TBC1D8	TBC1 domain family, member 8 (with GRAM domain)	-2,6	6,5E-03
ENSG00000153558	MRPS24	mitochondrial ribosomal protein S24	-2,3	1,2E-03
ENSG00000153898	NEU1	sialidase 1 (lysosomal sialidase)	-2,5	2,9E-02
ENSG00000154153	EVA1C	eva-1 homolog C (C. elegans)	-2,5	2,5E-02
ENSG00000154277	ANGPTL4	angiopoietin-like 4	-3,5	1,7E-02
ENSG00000155903	CLEC2B	C-type lectin domain family 2, member B	-3,1	9,6E-03
ENSG00000155961	RP11-73K9.3		-2,8	2,0E-02
ENSG00000155962	MLPH	melanophilin	-2,4	3,6E-02
ENSG00000156414	AFAP1L1	actin filament associated protein 1-like 1	-2,5	1,9E-03
ENSG00000157036	PIM3	pim-3 oncogene	-2,6	7,3E-03
ENSG00000158402	ADORA2B	adenosine A2b receptor	-2,6	7,2E-03
ENSG00000159167	PAQR5	progesterone and adipoQ receptor family member V	-2,9	9,1E-04
ENSG00000159433	ITPKA	inositol-trisphosphate 3-kinase A	-2,8	5,8E-03
ENSG00000159674	SLC43A3	solute carrier family 43, member 3	-2,6	3,7E-02
ENSG00000161217	TMEM38B	transmembrane protein 38B	-2,7	1,1E-02
ENSG00000162415	AC017104.6		-2,7	5,1E-03
ENSG00000162511	PMAIP1	phorbol-12-myristate-13-acetate-induced protein 1	-2,5	2,7E-03
ENSG00000162738	GAS2L3	growth arrest-specific 2 like 3	-2,5	1,6E-02
ENSG00000163132	UAP1L1	UDP-N-acetylglucosamine pyrophosphorylase 1-like 1	-3,2	2,1E-02
ENSG00000163879	FLVCR2	feline leukemia virus subgroup C cellular receptor family, member 2	-2,8	1,8E-02
ENSG00000164114	C15orf62	chromosome 15 open reading frame 62	-3,0	1,1E-02
ENSG00000164308	CPNE8	copine VIII	-2,6	4,1E-03
ENSG00000164530	EPOR	erythropoietin receptor	-2,5	5,8E-03
ENSG00000164542	TFAP2C	transcription factor AP-2 gamma (activating enhancer binding protein 2 gamma)	-3,0	4,3E-03
ENSG00000164692	THEMIS2	thymocyte selection associated family member 2	-3,2	1,4E-02
ENSG00000164742	TGFBR3	transforming growth factor, beta receptor III	-3,1	2,6E-04
ENSG00000165113	TCAP	titin-cap	-3,4	2,4E-02
ENSG00000167207	IFITM1	interferon induced transmembrane protein 1	-3,3	3,8E-02
ENSG00000168329	RHBDL3	rhomoid, veinlet-like 3 (Drosophila)	-3,2	6,9E-04
ENSG00000168961	TAS1R3	taste receptor, type 1, member 3	-3,7	2,1E-02
ENSG00000169946	COL28A1	collagen, type XXVIII, alpha 1	-3,5	5,4E-03
ENSG00000171084	ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1	-3,2	2,3E-05

Supplementary Data – Project 1

Table S1.4 (continued)

Ensembl gene ID	Gene name	Description	log2 FC	Adj p-value
Downregulated with log2FC>-1 and adjusted p-value <0,05				
ENSG00000171135	PLIN2	perilipin 2	-3,6	1,6E-03
ENSG00000172059	AP000240.9		-3,3	2,8E-02
ENSG00000172350	SOD2	superoxide dismutase 2, mitochondrial	-3,4	2,6E-02
ENSG00000173226	GS1-600G8.5		-3,7	7,8E-03
ENSG00000173960	RP11-720L2.3		-3,6	5,6E-03
ENSG00000175899	PTGES	prostaglandin E synthase	-3,4	1,4E-02
ENSG00000176406	ADRB2	adrenoceptor beta 2, surface	-3,0	1,1E-02
ENSG00000180828	SECTM1	secreted and transmembrane 1	-3,8	2,6E-03
ENSG00000182022	FOXL2	forkhead box L2	-2,8	8,3E-03
ENSG00000183856	CYP26B1	cytochrome P450, family 26, subfamily B, polypeptide 1	-4,3	5,6E-03
ENSG00000184408	THBS4	thrombospondin 4	-4,4	1,1E-02
ENSG00000184792	TRIM36	tripartite motif containing 36	-3,9	3,8E-02
ENSG00000185274	MYPN	myopalladin	-3,5	1,9E-02
ENSG00000185275	HIST1H2AE	histone cluster 1, H2ae	-3,8	3,3E-02
ENSG00000186281	TMEM26	transmembrane protein 26	-3,4	4,3E-06
ENSG00000186951	PGF	placental growth factor	-4,0	5,8E-03
ENSG00000187325	ATP8B4	ATPase, class I, type 8B, member 4	-3,6	6,1E-03
ENSG00000187764	NPTX1	neuronal pentraxin I	-5,4	1,8E-02
ENSG00000188313	RP11-134L10.1		-3,9	2,3E-04
ENSG00000189221	GRM1	glutamate receptor, metabotropic 1	-4,0	6,0E-03
ENSG00000196611	SLC7A14	solute carrier family 7, member 14	-4,3	1,6E-03
ENSG00000196923	TLR4	toll-like receptor 4	-3,9	1,2E-03
ENSG00000198399	TNFAIP6	tumor necrosis factor, alpha-induced protein 6	-4,0	1,1E-02
ENSG00000198948	SLCO5A1	solute carrier organic anion transporter family, member 5A1	-3,9	2,4E-02
ENSG00000203685	MAP7	microtubule-associated protein 7	-3,7	6,8E-03
ENSG00000204103	TFPI2	tissue factor pathway inhibitor 2	-4,8	2,0E-02
ENSG00000204287	HHEX	hematopoietically expressed homeobox	-4,1	1,3E-05
ENSG00000205078	OAS2	2'-5'-oligoadenylate synthetase 2, 69/71kDa	-4,6	3,1E-02
ENSG00000205246	RP11-930P14.2		-4,3	1,1E-02
ENSG00000214376	RP11-809M12.1		-4,6	2,3E-02
ENSG00000214900	RAB38	RAB38, member RAS oncogene family	-3,7	1,8E-02
ENSG00000224281	IL8	interleukin 8	-4,9	2,8E-02
ENSG00000225554	PDK4	pyruvate dehydrogenase kinase, isozyme 4	-5,2	1,3E-04
ENSG00000226471	SLC6A15	solute carrier family 6 (neutral amino acid transporter), member 15	-4,8	1,9E-02
ENSG00000231389	STC1	stanniocalcin 1	-5,4	2,3E-06
ENSG00000231528	CTD-2184D3.5		-5,9	1,9E-03
ENSG00000232472	GNA14	guanine nucleotide binding protein (G protein), alpha 14	-4,2	5,1E-03
ENSG00000234936	IL24	interleukin 24	-5,3	6,8E-03
ENSG00000235508	NEDD4L	neural precursor cell expressed, developmentally down-regulated 4-like, E3 ubiquitin protein ligase	-4,2	5,1E-03
ENSG00000236617	RP5-898J17.1		-5,9	2,6E-02
ENSG00000237515	FGD4	FYVE, RhoGEF and PH domain containing 4	-4,3	3,2E-03
ENSG00000237883	RP4-597N16.4		-5,0	6,1E-03
ENSG00000237927	FOXQ1	forkhead box Q1	-5,4	9,9E-05

Supplementary Data – Project 1

Table S1.4 (continued)

Ensembl gene ID	Gene name	Description	log2 FC	Adj p-value
Downregulated with log2FC > -1 and adjusted p-value < 0,05				
ENSG00000240476	CMPK2	cytidine monophosphate (UMP-CMP) kinase 2, mitochondrial	-5,7	1,1E-02
ENSG00000243137	VGF	VGF nerve growth factor inducible	-6,0	1,9E-02
ENSG00000243649	CD7	CD7 molecule	-5,4	4,6E-06
ENSG00000249790	BMP2	bone morphogenetic protein 2	-5,4	3,3E-03
ENSG00000250131	CXCL11	chemokine (C-X-C motif) ligand 11	-6,3	2,8E-02
ENSG00000251258	RP11-262H14.3		-7,0	1,0E-02
ENSG00000254510	FYB	FYN binding protein	-6,5	5,4E-05
ENSG00000255129	SCN1A	sodium channel, voltage-gated, type I, alpha subunit	-6,0	1,7E-02
ENSG00000257093	NYAP1	neuronal tyrosine-phosphorylated phosphoinositide-3-kinase adaptor 1	-5,5	2,4E-02
ENSG00000257315	RSAD2	radical S-adenosyl methionine domain containing 2	-6,6	1,2E-02
ENSG00000260083	HERC5	HECT and RLD domain containing E3 ubiquitin protein ligase 5	-5,5	4,9E-04
ENSG00000260336	WNT16	wingless-type MMTV integration site family, member 16	-4,8	1,5E-02
ENSG00000261625	AC104135.4		-4,8	2,4E-02
ENSG00000262003	KCNJ2	potassium inwardly-rectifying channel, subfamily J, member 2	-5,4	3,8E-04
ENSG00000262097	VIT	vitrin	-5,2	2,2E-02
ENSG00000263812	IFI27	interferon, alpha-inducible protein 27	-5,7	1,9E-02
ENSG00000264345	DLL4	delta-like 4 (Drosophila)	-6,3	3,6E-04
ENSG00000267365	LONRF3	LON peptidase N-terminal domain and ring finger 3	-7,2	4,6E-03
ENSG00000267669	ADH1B	alcohol dehydrogenase 1B (class I), beta polypeptide	-5,5	3,9E-04
ENSG00000268001	RP3-406A7.7		-5,7	1,9E-03
ENSG00000269821	EREG	epiregulin	-5,4	1,5E-02
ENSG00000270124	PPARGC1A	peroxisome proliferator-activated receptor gamma, coactivator 1 alpha	-6,4	7,2E-06
ENSG00000272189	RGPD2	RANBP2-like and GRIP domain containing 2	-6,1	2,3E-02
ENSG00000272476	NROB1	nuclear receptor subfamily 0, group B, member 1	-7,7	8,0E-06
ENSG00000273131	BST2	bone marrow stromal cell antigen 2	-6,9	1,1E-05
ENSG00000273356	SERPINB2	serpin peptidase inhibitor, clade B (ovalbumin), member 2	-6,2	6,7E-03
ENSG00000273387	LCNL1	lipocalin-like 1	-6,4	3,5E-03

**Table S1.5 Functional Annotation (-FCS/ ciliary conditions)**

<b>GO term: Biological Process</b>	<b>Count</b>	<b>%</b>	<b>PValue</b>	<b>Genes</b>	<b>List Total</b>	<b>Pop Hits</b>	<b>Pop Total</b>	<b>Fold Enrichment</b>	<b>Bonferroni</b>	<b>Benjamini</b>	<b>FDR</b>
<b>GO:0001649~osteoblast differentiation</b>	7	3,29E+15	1,11E-03	SHOX2, MSX2, BMP2, SEMA7A, GJA1, COL1A1, GLI2	189	104	16792	5,98E+14	0,80	0,41	1,82E+16
<b>GO:0030036~actin cytoskeleton organization</b>	7	3,29E+15	3,45E-03	GAS2L3, RND3, SHROOM4, TESK2, SH2B2, ITPKA, FGD4	189	130	16792	4,78E+15	0,99	0,71	5,55E+15
<b>GO:0006954~inflammatory response</b>	12	5,63E+15	3,64E-03	RPS6KA5, TNFAIP6, BMP2, TSPAN2, SEMA7A, CYP26B1, PSTPIP1, TNFRSF14, TLR4, IL24, CXCL11, THEMIS2	189	379	16792	2,81E+16	0,99	0,65	5,86E+14
<b>GO:0045892~negative regulation of transcription, DNA-templated</b>	14	6,57E+15	4,15E-03	MDFI, FOXL2, BMP2, KLF11, ARID5A, NROB1, GLI3, AHR, MSX2, RPS6KA5, HHEX, EREG, ID4, TWIST2	189	499	16792	2,49E+16	1,00	0,63	6,64E+15
<b>GO:0044344~cellular response to fibroblast growth factor stimulus</b>	4	1,88E+16	4,49E-03	CD44, DLL4, POSTN, COL1A1	189	30	16792	1,18E+16	1,00	0,60	7,17E+15
<b>GO:0007507~heart development</b>	8	3,76E+15	4,59E-03	SHOX2, BMP2, PLCE1, GJA1, EPOR, GLI2, GLI3, SOD2	189	183	16792	3,88E+15	1,00	0,56	7,32E+15
<b>GO:0001701~in utero embryonic development</b>	8	3,76E+15	5,16E-03	HHEX, BMP2, RRGRI1, GJA1, GLI2, GLI3, BCL2L11, MYH10	189	187	16792	3,80E+16	1,00	0,56	8,19E+15
<b>GO:0048167~regulation of synaptic plasticity</b>	4	1,88E+16	5,40E-03	LZTS1, SYNGAP1, VGF, ITPKA	189	32	16792	1,11E+16	1,00	0,54	8,56E+15
<b>GO:0007155~cell adhesion</b>	13	6,10E+15	5,65E-03	PCDH10, COL28A1, POSTN, CDH4, RND3, TNFAIP6, SORBS1, CD44, PSTPIP1, COL1A1, THEMIS2, MYH10, THBS4	189	459	16792	2,52E+16	1,00	0,52	8,94E+15
<b>GO:0060337~type I interferon signaling pathway</b>	5	2,35E+16	5,73E-03	IFI27, BST2, IFITM1, RSAD2, OAS2	189	64	16792	6,94E+16	1,00	0,50	9,07E+13
<b>GO:0000187~activation of MAPK activity</b>	6	2,82E+16	7,07E-03	BMP2, PLCE1, ADORA2B, TGFA, TLR4, GRM1	189	107	16792	4,98E+15	1,00	0,54	1,11E+16
<b>GO:0009409~response to cold</b>	4	1,88E+16	7,52E-03	ADRB2, VGF, PPARGC1A, SOD2	189	36	16792	9,87E+15	1,00	0,54	1,17E+16
<b>GO:0007267~cell-cell signaling</b>	9	4,23E+15	7,90E-03	TNFAIP6, BMP2, ADRB2, EREG, BST2, PGF, GJA1, TFAP2C, CXCL11	189	254	16792	3,15E+15	1,00	0,53	1,23E+15
<b>GO:0007512~adult heart development</b>	3	1,41E+16	8,97E-03	TCAP, GJA1, MYH10	189	13	16792	2,05E+14	1,00	0,56	1,38E+16
<b>GO:0060317~cardiac epithelial to mesenchymal transition</b>	3	1,41E+16	8,97E-03	WNT16, BMP2, TGFB3	189	13	16792	2,05E+14	1,00	0,56	1,38E+16
<b>GO:0010951~negative regulation of endopeptidase activity</b>	6	2,82E+16	1,17E-02	A2M, BST2, SERPINB7, SERPINB2, COL28A1, TFPI2	189	121	16792	4,41E+15	1,00	0,63	1,76E+15
<b>GO:0042060~wound healing</b>	5	2,35E+16	1,24E-02	EREG, SERPINB2, TGFA, IL24, GLI3	189	80	16792	5,55E+15	1,00	0,63	1,87E+16
<b>GO:0043434~response to peptide hormone</b>	4	1,88E+16	1,31E-02	HHEX, EREG, GJA1, COL1A1	189	44	16792	8,08E+15	1,00	0,63	1,95E+15

Table S1.5 (continued)

GO term: Biological Process	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0050731~positive regulation of peptidyl-tyrosine phosphorylation	5	2,35E+16	1,35E-02	CD44, HCLS1, TNFRSF14, RICTOR, THBS4	189	82	16792	5,42E+15	1,00	0,63	2,01E+16
GO:0008285~negative regulation of cell proliferation	11	5,16E+15	1,39E-02	MSX2, BMP2, EREG, IFITM1, PTGES, DLL4, HIST1H2AE, KLF11, IL24, GLI3, SOD2	189	396	16792	2,47E+15	1,00	0,62	2,07E+15
GO:0051781~positive regulation of cell division	4	1,88E+16	1,56E-02	EREG, PGF, TGFA, THBS4	189	47	16792	7,56E+15	1,00	0,64	2,29E+16
GO:0007157~heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules	4	1,88E+16	1,84E-02	TENM4, PTPRD, CD200, CDH4	189	50	16792	7,11E+15	1,00	0,69	2,65E+16
GO:0001501~skeletal system development	6	2,82E+16	1,90E-02	SHOX2, BMP2, ETS2, POSTN, COL1A1, GLI2	189	137	16792	3,89E+16	1,00	0,68	2,72E+16
GO:0006955~immune response	11	5,16E+15	2,03E-02	FYB, SECTM1, IL1R1, SEMA7A, TGFB3, TNFRSF14, TLR4, OAS2, IL24, CXCL11, CD7	189	421	16792	2,32E+15	1,00	0,69	2,87E+16
GO:0007264~small GTPase mediated signal transduction	8	3,76E+15	2,10E-02	RND3, RAB32, RALGPS2, PLCE1, RAB39B, RAB38, DOCK11, DOCK4	189	246	16792	2,89E+16	1,00	0,69	2,96E+15
GO:0042475~odontogenesis of dentin-containing tooth	4	1,88E+16	2,37E-02	BMP2, GLI2, GLI3, BCL2L11	189	55	16792	6,46E+16	1,00	0,72	3,27E+15
GO:0007173~epidermal growth factor receptor signaling pathway	4	1,88E+16	2,48E-02	RPS6KA5, PLCE1, EREG, TGFA	189	56	16792	6,35E+15	1,00	0,73	3,40E+14
GO:0009954~proximal/distal pattern formation	3	1,41E+16	2,93E-02	CYP26B1, GLI2, GLI3	189	24	16792	1,11E+16	1,00	0,77	3,88E+14
GO:0045732~positive regulation of protein catabolic process	4	1,88E+16	2,97E-02	CREBRF, DACT1, GJA1, NEDD4L	189	60	16792	5,92E+15	1,00	0,76	3,92E+16
GO:0043066~negative regulation of apoptotic process	11	5,16E+15	3,25E-02	MSX2, CD44, SERPINB2, TGFA, PIM3, GLI2, GLI3, FLNA, TWIST2, SOD2, ANGPTL4	189	455	16792	2,15E+16	1,00	0,78	4,21E+15
GO:0021776~smoothened signaling pathway involved in spinal cord motor neuron cell fate specification	2	9,39E-01	3,32E-02	GLI2, GLI3	189	3	16792	5,92E+15	1,00	0,78	4,28E+16
GO:0021775~smoothened signaling pathway involved in ventral spinal cord interneuron specification	2	9,39E-01	3,32E-02	GLI2, GLI3	189	3	16792	5,92E+15	1,00	0,78	4,28E+16
GO:0050727~regulation of inflammatory response	4	1,88E+16	3,36E-02	IL1R1, SEMA7A, RICTOR, ADAMTS12	189	63	16792	5,64E+15	1,00	0,78	4,32E+15
GO:0051607~defense response to virus	6	2,82E+16	3,82E-02	BST2, IFITM1, HERC5, RSAD2, OAS2, PMAIP1	189	165	16792	3,23E+16	1,00	0,81	4,75E+15



Table S1.5 (continued)

GO:0071222~cellular response to lipopolysaccharide	5	2,35E+16	3,82E-02	TLR4, IL24, ABCA1, PPARGC1A, CMPK2	189	113	16792	3,93E+15	1,00	0,80	4,75E+15
GO:0090200~positive regulation of release of cytochrome c from mitochondria	3	1,41E+16	3,90E-02	MLLT11, PMAIP1, BCL2L11	189	28	16792	9,52E+15	1,00	0,80	4,82E+15
GO:0001938~positive regulation of endothelial cell proliferation	4	1,88E+16	4,23E-02	BMP2, PGF, RICTOR, THBS4	189	69	16792	5,15E+15	1,00	0,81	5,10E+15
GO:0007442~hindgut morphogenesis	2	9,39E-01	4,40E-02	GLI2, GLI3	189	4	16792	4,44E+16	1,00	0,82	5,25E+15
GO:0060364~frontal suture morphogenesis	2	9,39E-01	4,40E-02	MSX2, GLI3	189	4	16792	4,44E+16	1,00	0,82	5,25E+15
GO:0045773~positive regulation of axon extension	3	1,41E+16	4,42E-02	SEMA7A, CDH4, ISLR2	189	30	16792	8,88E+15	1,00	0,81	5,26E+15
GO:0001666~response to hypoxia	6	2,82E+16	4,44E-02	BMP2, PGF, TGFBR3, POSTN, SOD2, ANGPTL4	189	172	16792	3,10E+16	1,00	0,81	5,28E+15
GO:0097190~apoptotic signaling pathway	4	1,88E+16	4,54E-02	IFI27, MLLT11, TNFRSF14, TLR4	189	71	16792	5,01E+15	1,00	0,80	5,36E+15
GO:0071260~cellular response to mechanical stimulus	4	1,88E+16	4,54E-02	GJA1, TLR4, COL1A1, KCNJ2	189	71	16792	5,01E+15	1,00	0,80	5,36E+15
GO:0045944~positive regulation of transcription from RNA polymerase II promoter	18	8,45E+15	4,89E-02	BMP2, FOXL2, MAFB, HCLS1, TLR4, GLI2, GLI3, PPARGC1A, AHR, RPS6KA5, SHOX2, HHEX, AHRR, ADRB2, DBP, ETS2, ID4, TFAP2C	189	981	16792	1,63E+16	1,00	0,82	5,64E+15
GO:0050873~brown fat cell differentiation	3	1,41E+16	4,97E-02	ADRB2, SH2B2, PPARGC1A	189	32	16792	8,33E+15	1,00	0,82	5,69E+15
GO:0035115~embryonic forelimb morphogenesis	3	1,41E+16	4,97E-02	SHOX2, MSX2, RPGRIP1L	189	32	16792	8,33E+15	1,00	0,82	5,69E+15
GO:0001889~liver development	4	1,88E+16	5,02E-02	DBP, RPGRIP1L, TGFBR3, SOD2	189	74	16792	4,80E+15	1,00	0,82	5,74E+16

Supplementary Data – Project 1

**Table S1.6 Alternative Splicing (+FCS)**

ID_IGV	Gene name	Description	log2 FC	Padj	dispersion
<b>JunctionSeq: log2FC &gt;3 or &lt;-3; Padj &lt;0,01; dispersion &lt; 0,05</b>					
chr19:35505290-35506729	GRAMD1A	GRAM domain containing 1A	3,05E+01	1,85E-25	4,76E-02
chr22:25855482-25855648	CRYBB2P1	crystallin, beta B2 pseudogene 1	3,19E+00	5,84E-35	1,45E-02
chr17:19642937-19642987	ALDH3A1	aldehyde dehydrogenase 3 family, member A1	-4,20E+00	1,69E-04	3,94E-02
chr17:19643648-19643683	ALDH3A1	aldehyde dehydrogenase 3 family, member A1	-4,09E+00	4,52E-05	3,01E-02
chr17:48265317-48266102	COL1A1	collagen, type I, alpha 1	1,07E+01	6,31E-52	4,70E-02
chr17:48266352-48267929	COL1A1	collagen, type I, alpha 1	3,20E+01	9,22E-48	3,23E-02
chr5:79355262-79355314	THBS4	thrombospondin 4	-2,58E+01	2,45E-03	4,12E-02
chr5:79375081-79375727	THBS4	thrombospondin 4	-3,72E+00	1,54E-05	8,75E-03
chr12:49521836-49521899	TUBA1B	tubulin, alpha 1b	-3,41E+01	2,48E-78	3,45E-02
chr1:120572528-120572586	NOTCH2	notch 2	3,32E+00	1,97E-32	1,97E-02
chr1:120572586-120572610	NOTCH2	notch 2	3,10E+00	1,15E-18	2,99E-02
chr1:120611947-120612240	NOTCH2	notch 2	3,65E+00	4,97E-34	2,35E-02
chr1:120548211-120572528	NOTCH2	notch 2	3,11E+00	1,48E-43	1,00E-02
chr1:120572610-120611947	NOTCH2	notch 2	3,03E+00	7,26E-15	3,41E-02
chr17:39978672-39978927	FKBP10	FK506 binding protein 10, 65 kDa	-3,17E+01	2,08E-84	1,55E-02
chr8:146015325-146015840	RPL8	ribosomal protein L8	-3,32E+01	1,38E-275	5,44E-03
chr1:228345440-228347527	GJC2	gap junction protein, gamma 2, 47kDa	4,50E+00	2,34E-03	3,50E-02
chr2:37010545-37014271	VIT	vitrin	-3,39E+00	7,66E-03	2,94E-02
chr3:49395131-49395501	GPX1	glutathione peroxidase 1	-3,14E+01	8,80E-105	9,00E-03
chr2:58479020-58479660	EIF3FP3	eukaryotic translation initiation factor 3, subunit F pseudogene 3	3,16E+00	4,36E-03	3,93E-02
ID_IGV	Gene name	Description	Padj	logef	deltapsi
<b>Leafcutter: p.adj.&lt;0,01; logef &gt;0,6 or &lt;-0,6; ΔΨ&lt;-0,5 or ΔΨ&gt;0,5</b>					
chr21:46708080-46708619:clu_17382	BX322557.10		7,83E-03	-9,36	-5,76E-01
chr1:110257862:110259959:clu_29147	GSTM1,GSTM5	glutathione S-transferase mu 1 ,glutathione S-transferase mu 5	9,26E-08	4,20	7,10E-01
chr17:49243858:49244188:clu_7988	NME1,NME1-NME2,NME2	NME/NM23 nucleoside diphosphate kinase 1 ,NME1-NME2 readthrough ,NME/NM23 nucleoside diphosphate kinase 2	4,28E-07	-3,84	-6,07E-01

Supplementary Data – Project 1

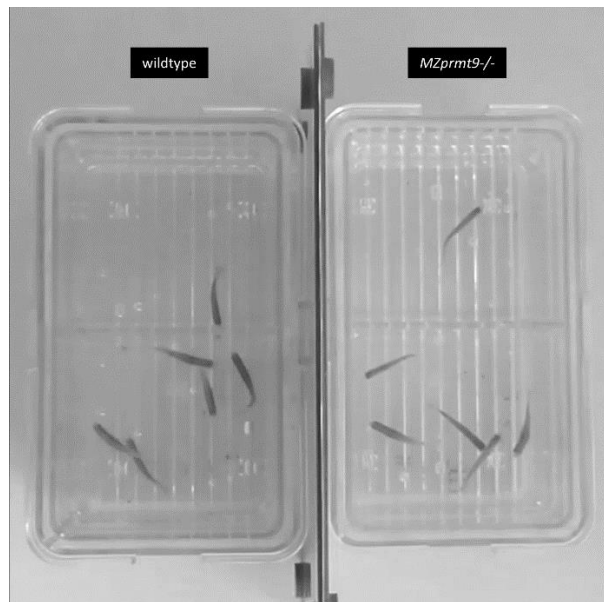
**Table S1.7 Alternative Splicing (-FCS/ciliary conditions)**

ID_IGV	Gene name	Description	log2 FC	Padj	dispersion
JunctionSeq: log2FC >3 or <-3; Padj <0,01; dispersion < 0,05					
chr17:48263333-48263677	COL1A1	collagen, type I, alpha 1	3,41E+01	0,00E+00	5,75E-03
chr11:6341550-6341571	PRKCDBP	protein kinase C, delta binding protein	3,16E+01	2,89E-83	1,75E-02
chrX:152966365-152967461	BCAP31	B-cell receptor-associated protein 31	3,09E+01	5,43E-25	4,67E-02
chr2:232319972-232320168	NCL	nucleolin	3,02E+01	1,63E-18	4,90E-02
chr19:35505290-35506729	GRAMD1A	GRAM domain containing 1A	3,00E+01	6,90E-36	2,96E-02
chr9:132481697-132482886	PRRX2	paired related homeobox 2	2,98E+01	5,02E-22	3,45E-02
chr7:73471043-73471720	ELN	elastin	2,96E+01	1,99E-21	2,38E-02
chr2:216293012-216296644	FN1	fibronectin 1	2,94E+01	4,97E-22	4,71E-02
chr17:48265344-48265890	COL1A1	collagen, type I, alpha 1	2,74E+01	8,63E-13	2,58E-02
chr11:66082958-66083102	CD248	CD248 molecule, endosialin	9,90E+00	2,69E-114	8,57E-03
chr4:174312579-174320218	SCRG1	stimulator of chondrogenesis 1	5,20E+00	4,19E-17	2,00E-02
chr4:174320218-174320394	SCRG1	stimulator of chondrogenesis 1	5,01E+00	4,35E-34	1,31E-02
chr7:73471711-73471798	ELN	elastin	4,94E+00	3,27E-18	2,46E-02
chr1:120572586-120572610	NOTCH2	notch 2	3,94E+00	7,89E-17	4,97E-02
chr1:120572610-120611947	NOTCH2	notch 2	3,78E+00	2,45E-18	4,15E-02
chrX:135279184-135279221	FHL1	four and a half LIM domains 1	3,04E+00	2,82E-09	4,02E-02
chr8:22421819-22421875	SORBS3	sorbin and SH3 domain containing 3	- 3,09E+00	4,69E-65	8,01E-03
chr8:22409207-22409439	SORBS3	sorbin and SH3 domain containing 3	- 3,11E+00	1,12E-09	4,90E-02
chr8:22421590-22421795	SORBS3	sorbin and SH3 domain containing 3	- 3,14E+00	5,85E-35	1,49E-02
chr8:22421795-22421819	SORBS3	sorbin and SH3 domain containing 3	- 3,16E+00	8,83E-39	1,44E-02
chr8:22421882-22421980	SORBS3	sorbin and SH3 domain containing 3	- 3,17E+00	1,50E-61	8,12E-03
chr8:22421875-22421882	SORBS3	sorbin and SH3 domain containing 3	- 3,18E+00	1,31E-48	1,11E-02
chr8:22412410-22412420	SORBS3	sorbin and SH3 domain containing 3	- 3,20E+00	1,88E-18	3,28E-02
chr3:54880419-54880494	CACNA2D3	calcium channel, voltage-dependent, alpha 2/delta subunit 3	- 3,21E+00	2,84E-04	4,45E-02
chr8:22409439-22409535	SORBS3	sorbin and SH3 domain containing 3	- 3,22E+00	1,63E-18	3,21E-02
chr8:22418847-22418886	SORBS3	sorbin and SH3 domain containing 3	- 3,23E+00	2,50E-28	2,18E-02
chr8:22418886-22419377	SORBS3	sorbin and SH3 domain containing 3	- 3,23E+00	1,87E-32	1,74E-02
chr8:22412420-22412493	SORBS3	sorbin and SH3 domain containing 3	- 3,27E+00	3,79E-33	2,00E-02
chr8:22421980-22422034	SORBS3	sorbin and SH3 domain containing 3	- 3,28E+00	4,89E-80	7,26E-03
chr8:22412375-22412410	SORBS3	sorbin and SH3 domain containing 3	- 3,28E+00	1,85E-17	3,52E-02
chr8:22414421-22415641	SORBS3	sorbin and SH3 domain containing 3	- 3,32E+00	1,00E-29	2,11E-02
chr8:22415705-22418847	SORBS3	sorbin and SH3 domain containing 3	- 3,32E+00	9,28E-21	3,14E-02
chr8:22414418-22414421	SORBS3	sorbin and SH3 domain containing 3	- 3,35E+00	2,31E-26	2,48E-02
chr8:22409535-22411964	SORBS3	sorbin and SH3 domain containing 3	- 3,35E+00	1,99E-21	2,79E-02
chr8:22421499-22421590	SORBS3	sorbin and SH3 domain containing 3	- 3,37E+00	9,30E-83	7,46E-03

Supplementary Data – Project 1

Table S1.7 (continued)

ID_IGV	Gene name	Description	log2 FC	Padj	dispersion
<b>JunctionSeq: log2FC &gt;3 or &lt;-3; Padj &lt;0,01; dispersion &lt; 0,05</b>					
chr8:22411964-22412121	SORBS3	sorbin and SH3 domain containing 3	- 3,38E+00	1,89E-20	3,43E-02
chr8:22422034-22422780	SORBS3	sorbin and SH3 domain containing 3	- 3,39E+00	1,87E-90	5,39E-03
chr8:22412493-22414227	SORBS3	sorbin and SH3 domain containing 3	- 3,39E+00	6,20E-31	2,12E-02
chr8:22415641-22415705	SORBS3	sorbin and SH3 domain containing 3	- 3,39E+00	2,96E-31	2,26E-02
chr8:22419377-22419444	SORBS3	sorbin and SH3 domain containing 3	- 3,40E+00	2,49E-54	1,22E-02
chr8:22419444-22421499	SORBS3	sorbin and SH3 domain containing 3	- 3,42E+00	1,17E-84	5,61E-03
chr1:203149261-203149597	CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)	- 3,56E+00	3,31E-07	3,98E-02
chr8:22414227-22414418	SORBS3	sorbin and SH3 domain containing 3	- 3,58E+00	1,53E-39	2,04E-02
chr1:203149125-203149261	CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)	- 4,27E+00	1,07E-07	4,51E-02
chr3:170150410-170150569	CLDN11	claudin 11	- 2,88E+01	3,57E-21	3,80E-02
chr8:145059157-145059238	PARP10	poly (ADP-ribose) polymerase family, member 10	- 2,93E+01	1,05E-16	4,72E-02
chr3:13659785-13660420	FBLN2	fibulin 2	- 2,94E+01	2,96E-21	3,99E-02
chr19:17622580-17622675	PGLS	6-phosphogluconolactonase	- 3,06E+01	2,69E-31	3,66E-02
chr5:149494845-149495009	PDGFRB	platelet-derived growth factor receptor, beta polypeptide	- 3,10E+01	5,10E-84	1,30E-02
ID_IGV	Gene name	Description	Padj	logef	deltapsi
<b>Leafcutter: p.adj.&lt;0,01; logef &gt;0,6 or &lt;-0,6; ΔΨ&lt;-0,5 or ΔΨ&gt;0,5</b>					
chr2:178188124:178199121:clu_26639	AC074286.1,NFE2L2	nuclear factor, erythroid 2-like 2	6,33E-03	- 1,12E+01	-8,07E-01
chr11:76164432:76169227:clu_4745	C11orf30	chromosome 11 open reading frame 30	4,22E-03	1,93E+00	5,56E-01
chr7:14028764:14029188:clu_17671	ETV1	ets variant 1	5,46E-03	- 1,57E+00	-5,10E-01
chr1:110257862:110259959:clu_29147	GSTM1,GSTM5	glutathione S-transferase mu 1 ,glutathione S-transferase mu 5	6,49E-06	3,25E+00	6,75E-01
chr12:52636921:52639196:clu_2139	KRT7	keratin 7	8,32E-06	- 2,79E+00	-5,88E-01
chr17:41459028:41465657:clu_7691	LINC00910	long intergenic non-protein coding RNA 910	3,82E-04	- 4,59E+00	-7,61E-01
chr17:49243858:49244188:clu_7988	NME1,NME1-NME2,NME2	NME/NM23 nucleoside diphosphate kinase 1 ,NME1-NME2 readthrough ,NME/NM23 nucleoside diphosphate kinase 2	3,52E-07	- 4,43E+00	-6,04E-01
chr4:174312579:174387055:clu_23109	SCRG1	stimulator of chondrogenesis 1	2,47E-03	4,22E+00	9,55E-01
chr4:174312579:174320219:clu_23109	SCRG1	stimulator of chondrogenesis 1	2,47E-03	- 4,22E+00	-9,55E-01
chr5:155108550:15550866:clu_21796	SGCD	sarcoglycan, delta (35kDa dystrophin-associated glycoprotein)	1,67E-03	9,28E+00	5,41E-01
chr6:170062520:170064262:clu_20746	WDR27	WD repeat domain 27	4,06E-04	- 2,72E+00	-5,84E-01



**Movie S1. *MZprmt9*<sup>-/-</sup> mutants display behavioral anomalies.** To assess the behavior of adult zebrafish mutants a social behavior test was performed. The spatial segregation of a single fish from its group, in a tank that is divided only by a transparent disc, in general leads to the fact that the single fish stays most of the time close to its conspecific group. While the percentage of the time spend in the conspecific sector did not defer between *MZprmt9*<sup>-/-</sup> and wildtype animals, *MZprmt9*<sup>-/-</sup> mutants moved in general significantly slower.

### Authors contributions

S.S., E.S., K.K., K.C., W.I., S.T. R.C., O.B., T.K., S.F., A.E., L.C., W.S., R.J.,A.H., N.H. and H.D. gathered data from patients and performed clinical investigations; J.M. and V.G. gathered WES sequencing data and performed analyses; P.D., **A.KH.**, J.M., A.P. and F.M. gathered RNA-Seq data and performed analyses, **A.KH.**, M.P., A.S and C.S performed cell biology experiments and data analyses; **A.KH.**, C.E. and U.S. designed and performed the zebrafish experiments and data analyses; **A. KH.** and J.M. analyzed the data and wrote the paper; C.E., C.S. and U.S. contributed to manuscript writing; U.S. and H.D provided direction for the project, conceived and designed the experiments.

# Project 2

Based on:

Article was submitted to the Journal of Medical Genetics:

## ***MYH10* dominant variant causes Baraitser-Winter cerebro-fronto-facial syndrome related phenotype with major ophthalmic developmental features**

*Ariane Kröll-Hermi*<sup>1,2\*</sup>, *Sophie Scheidecker*<sup>1,3\*</sup>, *Séverine Bär*<sup>4\*</sup>, *Jean Muller*<sup>1,3</sup>, *Véronique Geoffroy*<sup>1</sup>, *Elise Schaefer*<sup>1,5</sup>, *Stéphane Kremer*<sup>6</sup>, *Christelle Etard*<sup>2</sup>, *Corinne Stoetzel*<sup>1</sup>, *Uwe Strähle*<sup>2</sup>, *Sylvie Friant*<sup>#</sup> and *Hélène Dollfus*<sup>1,7#</sup>

\* joint first authors

# equal contributors

1. Laboratoire de Génétique médicale, UMR\_S INSERM U1112, IGMA, Faculté de Médecine FMTS, Université de Strasbourg, Strasbourg, France.
2. Institut für Toxikologie und Genetik, Campus Nord, Karlsruher Institut für Technologie, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein Leopoldshafen, Germany.
3. Laboratoires de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France.
4. Laboratoire de Génétique Moléculaire, Génomique, Microbiologie (GMGM), UMR7156, Centre National de Recherche Scientifique (CNRS), Université de Strasbourg, 67084 Strasbourg, France.
5. Service de Génétique Médicale, Hôpitaux Universitaires de Strasbourg, Strasbourg, France
6. Service de Neuroradiologie/Imagerie 2, CHU de Strasbourg, Hôpital de Hautepierre, Strasbourg, France.
7. Centre de Référence pour les affections rares en génétique ophtalmologique, CARGO, Filière SENSGENE, Hôpitaux Universitaires de Strasbourg, 67091 Strasbourg, France.

## 4.2 Project 2

Parts of this chapter (4.2.2-4.2.5) were reproduced with permission of the authors from the article: **“MYH10 dominant variant causes Baraitser-Winter cerebro-fronto-facial syndrome related phenotype with major ophthalmic developmental features”** that was submitted to the **“Journal of Medical Genetics”**. The synopsis will give a brief summary about the project and points out my personal contribution to this work.

### 4.2.1 Synopsis

Worldwide, only one out of 100,000 individuals is affected by coloboma.<sup>81</sup> Coloboma is usually caused by a failed optic fissure closure during the 5th to 7th week of fetal life.<sup>80</sup> The combination of coloboma and periorbital anomalies is even more rare and was only observed in very few syndromes, such as the Baraitser Winter Cerebro Fronto Facial syndrome (BWCF).<sup>270</sup>

A combination of coloboma, hypertelorism, ptosis and epicanthus inversus was also observed in one of the families we were working on. Using WES, we were not able to identify a mutation in one of the BWCF associated genes *ACTB* or *ACTG1*, but identified a heterozygous missense mutation in the *MYH10* gene (p.[Arg1471Pro];[=]) in all three affected family members. *MYH10* encodes for a non-muscular myosin and is an actin-dependent motor protein and interestingly, is supposed to interact with *ACTB* and *ACTG1* (FunCoup 4.0 human database).<sup>271</sup>

In order to investigate the impact of the heterozygous missense mutation on the protein function, Western Blots and immunostainings were conducted showing a diminished MYH10 protein level in patient's cells compared to control cells. Moreover, MYH10 was shown to co-localize with actin fibers and revealed that actin fibers in patient's cells are shorter and disorganized compared to actin fibers in control cells.

To get a better understanding of the gene function and to prove the link between *myh10* deficiency and the patients' phenotype, I first created a zebrafish knockout mutant for *myh10*. *myh10*<sup>-/-</sup> mutants are not lethal, survive until adulthood and become sexually mature. Surprisingly, no obvious anomalies were observed. *myh10*<sup>-/-</sup> zebrafish embryos develop normally and do not seem to display an ocular phenotype suggesting genetic compensation likely induced by nonsense mediated decay. To avoid genetic compensation, a second strategy was used to reduce the expression of *myh10*: the knock down *via* morpholino. A start-site morpholino as well as a splice-site morpholino were injected into one-cell stage embryos and resulted in both cases in a strong muscle phenotype with unorganized muscle fibers and a curved body axis. A very interesting result, as MYH10 has already been suggested previously to be an important component of the pre-myofibrilla in zebrafish.<sup>272</sup> Myh10 therefore seems to be important in the early development of zebrafish skeletal muscle, most likely in association with actin.

In addition, morphants have a smaller eye size and the optic fissure closes with one day delay. A distinct coloboma phenotype, however, was not observed.

In conclusion, in project 2, we report the identification of a novel heterozygous missense mutation identified in the gene *MYH10*, which leads to a BWCF-like syndrome. Moreover, we show that *myh10* deficiency causes an abnormal actin length and organization in patient's cells and affects the eye and muscle development in zebrafish.

### Personal contribution

To ensure that no possible candidate genes were overlooked or accidentally excluded from the analysis, the NGS data in our laboratory are always analyzed from at least two people. For project 2, I carried out the variant filtration for a second time and identified, as my colleague Corinne Stoetzel *myh10* as the potentially most interesting candidate gene. My main task for project 2, however, was the generation of a zebrafish model to prove the link between *MYH10* deficiency and the ocular phenotype observed in human patients (Figure 19). Moreover, I wrote main parts of the manuscript. For a detailed description of the authors contributions see the supplemental Information of Project 2 (S2).

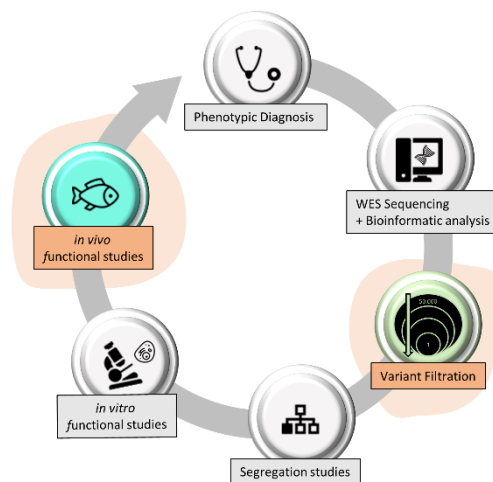


Figure 19. Personal contribution to project 2. Icons were created with icons8.com.



#### 4.2.2 Introduction – Project 2

Developmental anomalies of the eyeball may occur in isolation or in a syndromic context that may be associated to other developmental features with often a genetic origin.<sup>273,274</sup> Orbito palpebral developmental anomalies (around the eyeball) are also well described, possibly of genetic origin, and include a wide range of features ranging from craniofacial bone anomalies, such as: hypertelorism, to soft tissue abnormalities of various origins such as eyelid anomalies or congenital ptosis. However, syndromes combining both eyeball developmental anomalies and periocular developmental anomalies have been seldom described except for a few entities mostly associating coloboma and ptosis such as Noonan syndromes (MIM # 163950; MIM # 605275; MIM # 609942; MIM # 610733; MIM # 611553; MIM # 613224; MIM # 613706; MIM # 615355; MIM # 616559; MIM # 616564) or Baraitser-Winter cerebro-fronto-facial syndromes (BWCF) (MIM # 243310; MIM # 614583) belonging to actinopathies a novel group of conditions defined by their common molecular context.

We report herein on a family in which the mother and her two daughters presented with the association of eyeball developmental anomalies and congenital ptosis among other variable craniofacial features. The diagnosis of BWCF syndrome was considered but no pathogenic variants in the actin *ACTB* and *ACTG1* genes were detected. To elucidate the underlying molecular mechanism, we performed Whole Exome Sequencing (WES) and identified a likely pathogenic missense variant at the heterozygous state in the *MYH10* gene encoding the non-muscle myosin II heavy chain protein B (NMHC-II). The actin motor proteins NMHC-II have a central role in several cellular processes, as well as during embryogenesis and organogenesis. *MYH10* enriched in neuronal tissues during mouse embryonic development is involved in reorganization of the actin cytoskeleton and interacts *in silico* with *ACTB* and *ACTG1*, two BWCF genes.<sup>275,276</sup> Therefore, functional analyses using patient cells and zebrafish models were performed to explore the phenotypes due to *MYH10* loss-of-function mutations, showing its requirement for eye development in association with the actin *ACTB* and *ACTG1* genes implied in BWCF syndromes.

## 4.2.3 Results – Project 2

**Patients presenting with major ophthalmic features**

The mother and her two daughters (born from different fathers), had in common eyeball malformations as well as congenital ptosis and facial dysmorphism features including highly arched eyebrows, broad nasal bridge and long philtrum (Table 16 for clinical summary) (Figure 20 for main clinical features, Figure 21A for the pedigree of the family).

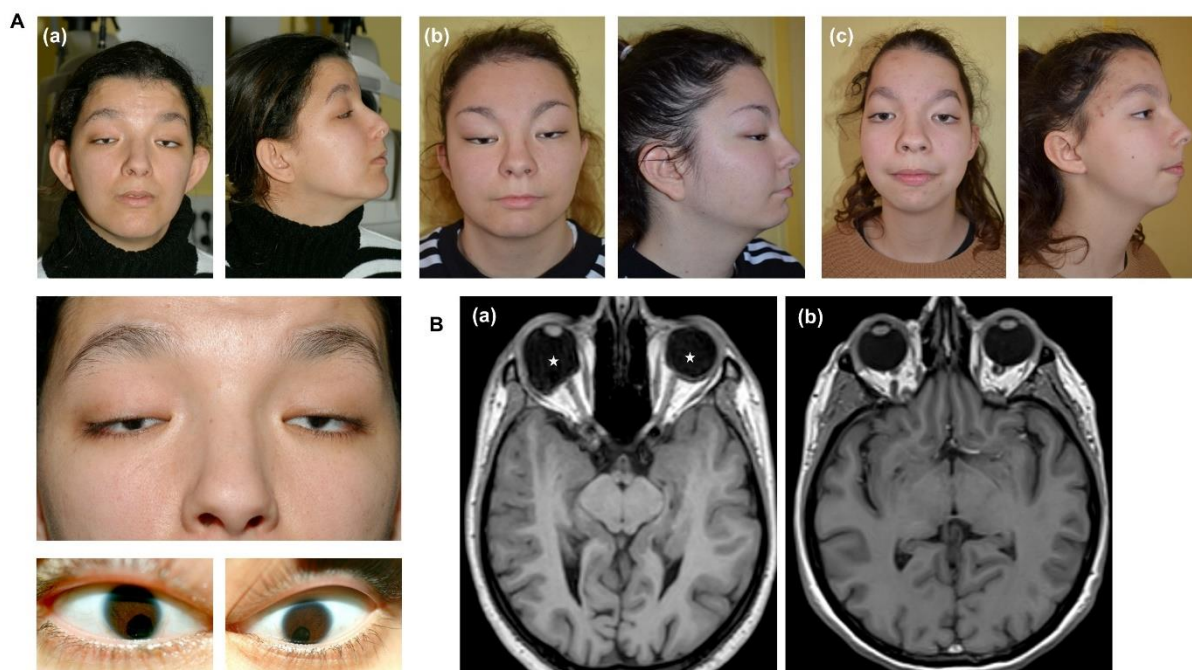
**Table 16. Main dysmorphic features of the 3 affected individuals and observed in actinopathies.<sup>277</sup>**

	II.2	III.1	III.2	BWCFE reported clinical manifestations
<b>Eye anomalies</b>				
Highly arched eyebrows	+	+	+	+
Congenital ptosis	+	+	+	+
Hypertelorism	+	+	+	+
Epicanthus inversus	+	+	+	+
Iridochorioretinal coloboma	+	+	-	+
Microcornea	+	-	-	
Oculomotor limitation	+	+	+	
<b>Other facial dysmorphism features</b>				
<b>Nose:</b>				
short, upturned with broad nasal bridge	+	+	+	+ (85 %)
anteverted nares, thick columella	+	+	+	+
Philtrum: long and smooth	+	+	+	+ (84 %)
<b>Mouth:</b>				
thin vermilion border of the upper lip, wide mouth, everted vermilion of the lower lip	+	+	+	+
cleft lip and palate (10 %)	-	-	-	+
Malar region: retrusion	+	+	+	+
Chin: retrognathia, pointed chin	+	+	+	+
Ears: abnormal shape, protruding	+	+	+	+ (73 %)
Prominent metopic ridge or trigonocephaly	-	-	+	+ (65 %)

The mother (individual II.2) presented with bilateral microcornea and iridochorioretinal coloboma. She presented also major bilateral ptosis as well as epicanthus inversus. Her first daughter (III.1) presented also a major bilateral ptosis and bilateral epicanthus inversus as well as iridal and retinal colobomas bilaterally. The second daughter (III.2) presented also with bilateral ptosis and epicanthus inversus. Ophthalmologic examination showed abnormal pigment epithelium anomalies at the level of the retina. Additionally, she had a history of trigonocephaly.

None of the patients had a history of learning disabilities or intellectual deficiency and cerebral MRI performed on the 3 patients did not reveal significant anomalies observed in BWCFE syndrome. No other family history was noted except for the grandfather (I.2) who was reported by members of the family to present with a ptosis and similar but milder facial gestalt. He could not be examined in our center, however an ID photograph confirmed the report of the family, as we could observe ptosis,

highly arched eyebrows, broad nasal bridge, long philtrum, thin upper lip, pointed chin and enlarged ears.



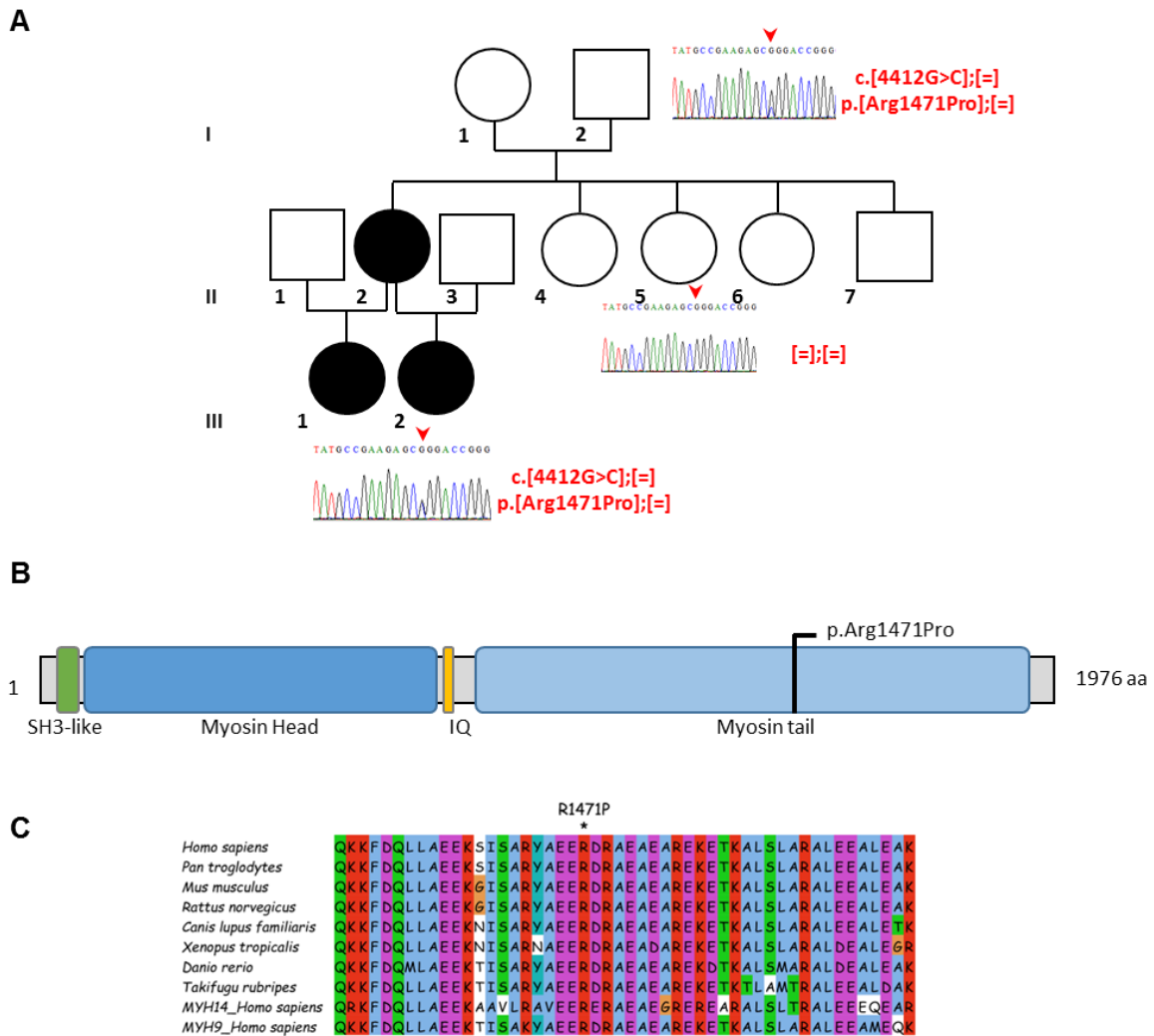
**Figure 20. Pictures of the patients.**

**(A)** Pictures of the patients showing the eye anomalies (highly arched eyebrows, ptosis, hypertelorism, epicanthus inversus in the 3 individuals and iridal coloboma for II.2) and close up photograph showing bilateral iridal coloboma with microcornea for II.2) and facial dysmorphism features of the affected individuals II.2 (a), III.1 (b) and III.2 (c). **(B)** Cerebral Magnetic Resonance Imaging. (a) Axial T1 weighted MR image showing bilateral eye ball morphological abnormality (stars) in patient II.2 compared to a non-affected individual (b).

### Whole exome sequencing identifies a heterozygous variant in the *MYH10* gene in the family

To identify causative variants, whole exome sequencing (WES) was performed for the two affected siblings (III.1, III.2) and the affected mother (II.2). From 100 754 to 103 655 genetic variants (SNV/indel/SV) were identified per proband. No pathogenic variants were detected in *ACTB* and *ACTG1*. Given the clear autosomal dominant inheritance and a candidate gene approach based on the BWCF syndrome hypothesis (Method S2.1), bioinformatic analyses narrowed down the number of variants to 5 potentially disease-causing heterozygous variations. Finally, 4 variants were manually ruled out based on gene function or familial segregation leading to the identification of a unique heterozygous missense variant in exon 32 of *MYH10* gene (c.4412G>C, p.Arg1471Pro). The Arginine at the 1471 position is conserved through the evolution and is localized in the coiled-coil tail domain (Figure 21A and B). This variant is predicted to be pathogenic by SIFT<sup>278</sup> and PolyPhen-2 (PPH2)<sup>279</sup> and is absent from any variations database (e.g. gnomAD<sup>182</sup>). Interestingly, the segregation study by Sanger sequencing suggested a mosaic status in individual I.2 who presented with a milder phenotype according to the ID photographs. The variant was not detected in a healthy sister of individual II.2

(Figure 21A). Samples of other members of the family were not available. To date, no additional patients carrying a *MYH10* variant with a similar phenotype could be identified using GeneMatcher<sup>172</sup> or by contacting the ERN-EYE network and other ophthalmic geneticists worldwide.



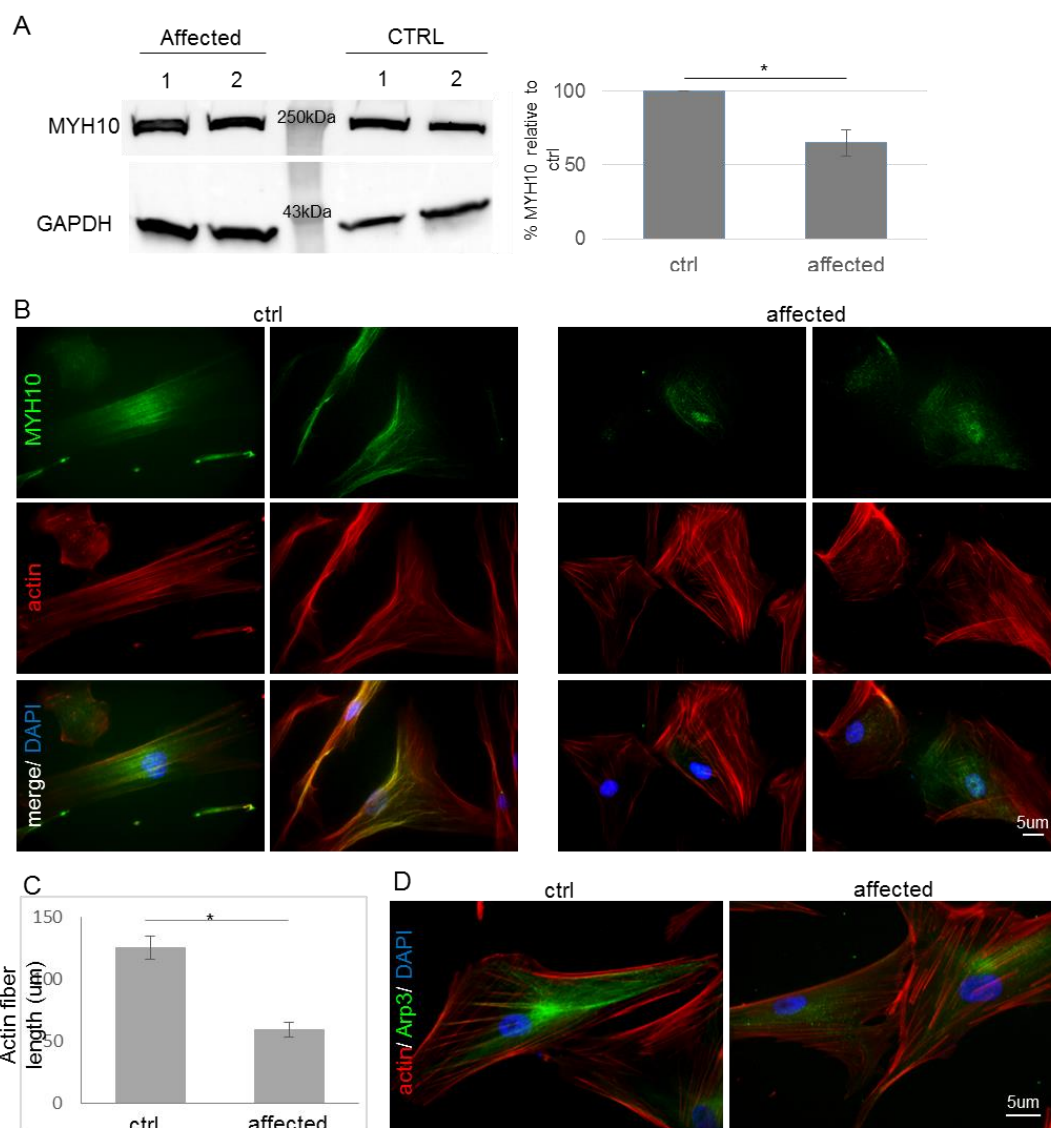
**Figure 21. Three affected family members carry a heterozygous *MYH10* variant.**

**(A)** Pedigree of the family with the segregation of the heterozygous *MYH10* variant (NM\_005964.3:c.4412G>C) identified in the 3 affected individuals (II.2, III.1 and III.2) and not detected in an unaffected sister II.5. The electropherogram in the mildly affected grandfather I.2 suggests a mosaic status. **(B)** Schematic representation of *MYH10* with the protein domains described according to the SMART and PFAM databases as follows: Myosin N-terminal SH3-like domain (33-73), IQ (787-807), Myosin\_head (87-771), Myosin\_tail\_1 (848-1928). **(C)** Multiple protein sequence alignment extracted from KOG0161 from the eggNOG 4.5.1 database and displayed thanks to the Jalview 2.11 program.

**MYH10 is less expressed and mis-localized in primary patient fibroblasts.**

To determine the cellular levels of *MYH10* in the patient’s primary fibroblasts, we assessed both the RNA and protein levels compared to non-affected controls. The intracellular level of *MYH10* RNA is similar in the patient and the controls. Whereas the *MYH10* protein level is significantly decreased in the patient fibroblasts compared to control cells (Figure 22A). These results suggest that the missense variant localized in the myosin tail - coil-coiled domain could affect the stability of the *MYH10* protein.

Since the MYH10 tail domain is required for assembly of the myosin II filaments<sup>280</sup>, we investigated whether the localization of the MYH10 protein could be disturbed. Therefore, we performed an MYH10 immunofluorescence assay on patient and non-affected primary fibroblasts (Figure 22B). MYH10 was mis-localized in the patient cells compared to the control, indeed the protein was distributed throughout the cell and the filamentous staining was not observed.



**Figure 22. MYH10 is less expressed and mis-localized in affected fibroblasts than control (ctrl) cells, affecting actin filaments length and cytoskeleton dynamics.**

**(A)** Fibroblasts from two control individuals (ctrl) and from the patient (two separate batches of cells) were collected and MYH10 detected by western blot in the whole cell lysate. GAPDH was used as a loading control and quantification of MYH10 relative to GAPDH performed using Image Lab. Data shown represent the mean of three independent experiments, involving cells from two to five different control individuals and two to three batches of patient cells. \*:p<0,05. **(B)** and **(C)** Cells were grown on coverslips, fixed, permeabilized, actin stained with Rhodamine-phalloidin, MYH10 **(B)** or Arp3 **(D)** labeled with the appropriate antibody and nuclei stained with DAPI. The actin-MYH10 **(B)** and actin-Arp3 **(D)** colocalization were observed in confocal microscopy.

### Alteration of the actin cytoskeleton structure and dynamics in affected fibroblasts.

Since MYH10 is known to be involved in the restructuring of the actin cytoskeleton and therefore plays a role in cell adhesion, cell migration and tissue architecture, we stained the actin filaments by

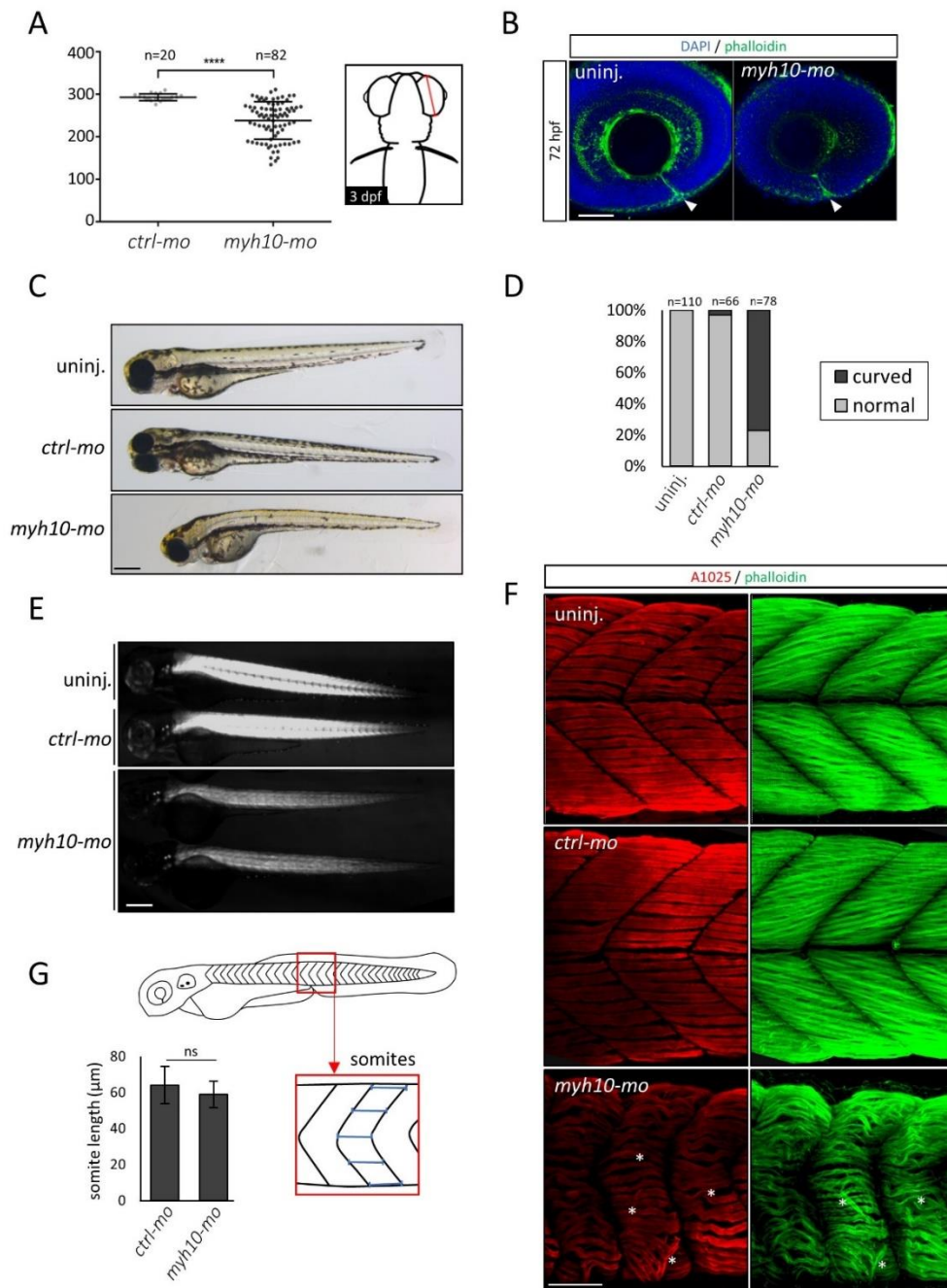
using Rhodamine-phalloidin in the different cells. A lack of colocalization between MYH10 and actin filaments was observed in the patient cells compared to the control (Figure 22B). Then, as non-muscle myosin II has been shown to play a direct role in the actin network assembly/disassembly dynamics<sup>281</sup>, actin filaments in patient primary fibroblasts were compared to those of non-affected individuals. After co-staining with DAPI, observation on a fluorescence microscope showed that actin fibers appear strikingly shorter in affected cells compared to the control (Figures 22B and C, Supplemental Figure S2.1). Indeed, in the patient's cells we observed a mix of actin fibers of different lengths showing massive disorganization. To quantify this observation, we determined the length of 15 to 20 fibers in the patient and control fibroblasts in five independent experiments. As shown in Figure 22C and Supplemental Figure S2.1, on average, actin fibers in the fibroblasts from the patient were about half as long as those in the non-affected control cells. This suggests that the *MYH10* variant present in the patient cells affects the actin network dynamics.

Then we investigated the colocalization of actin filaments with Arp3, one of the effectors of the Arp2/3 complex involved in actin polymerization and branching. In patient cells, Arp3 was redistributed into the cytoplasm and no longer localized on actin filaments (Figure 22D), indicating abnormal actin cytoskeleton dynamics in the patient cells.

#### ***myh10* zebrafish morphants present eye anomalies and impaired muscle development.**

As a link between *myh10* deficiency in zebrafish with an abnormal eye morphology has been previously suggested but not further investigated<sup>282</sup>, we examined the role of Myh10 during the optic fissure closure. The zebrafish orthologue (Ensembl ENSDARG0000000103.12) of human *MYH10* is located on LG 6 and has three predicted isoforms (1992 aa, 1986 aa and 1986 aa), all starting with the same start codon and sharing 78 % sequence identity with the longest human isoform (ENST00000360416.7). Injection of morpholino targeting the start site of *myh10* (*myh10-mo*) resulted in significant smaller eyes in *myh10* morphants at 3 dpf (Figure 24A). *myh10* morphants did not present a typical coloboma phenotype with a wide-open and non-closing optic cup (Figure 24B). However, a slight delay of one day between two- and three-days postfertilization of the optic fissure fusion process was observed while performing an *in situ* analysis on 24, 48 and 72 hpf staged embryos using a probe targeting *pax2a*, a gene related to coloboma and the optic fissure closure (supplemental Figure S2.2). Moreover, we observed a curved body axis in 78 % of *myh10* morphants (Figure 24C and D). Measure of birefringence at 3 dpf indicated an abnormal muscle integrity in *myh10* morphants (Figure 24E, F and G). Immunohistology against myosin and F-actin revealed that the striation was unaffected, however myofibrils are not stretched between myosepta leading to a wavy appearance with some fibers detached from the myoseptum. Using a second morpholino, targeting a splice site of *myh10*, we were





**Figure 23. Knock-down of *myh10* affects the eye development and the organization of skeletal muscle fibers in 3-day-old zebrafish embryos.**

**(A)** The diameter of the eye of 3-day-old *myh10* morphants are smaller compared to the eyes of control injected embryos. Statistical significance was determined using the Mann-Whitney Test, ns = non-significant. \*  $p < 0.05$ . \*\*  $p < 0.01$ . \*\*\*  $p < 0.001$ . \*\*\*\*  $p < 0.0001$ . **(B)** 3-day-old embryos were immunostained with phalloidin (in green) and DAPI (in blue) to visualize the optic fissure closure. The optic fissure of both, uninjected and *myh10-mo* injected embryos, was closed. **(C)** Representative image of 3-days-old uninjected, *ctrl-mo* injected and *myh10-mo* injected embryos. Scale bar 250 µm. **(D)** Quantification of embryos with a curved body axis. **(E)** Compared to uninjected and *ctrl-mo* injected embryos, *myh10* morphants display reduced birefringence indicating an impaired muscle organization. Scale bar 50 µm. **(F)** Immunostaining, with antibodies against myosin (A1025, red) and phalloidin toxin (green) confirmed the suspicion of an impaired muscle structure in *myh10-mo* injected embryos. Slow and fast muscle fibers in *myh10* morphants were clearly disorganized. White asterisks indicate fibers, that were detached from the myoseptum. Scale bar 50 µm. **(G)** Measuring the somite sizes, no significant difference was detectable. Statistical significance was determined using the unpaired t-Test, ns = non-significant. \*  $p < 0.05$ . \*\*  $p < 0.01$ . \*\*\*  $p < 0.001$ . \*\*\*\*  $p < 0.0001$ .

able to reproduce the eye and muscle phenotype (supplemental Figure S2.4). All together the zebrafish data indicate a role of *myh10* during the eye and muscle development.

In addition, a zebrafish knockout mutant for *myh10* was generated (*myh10*<sup>ka708</sup>) to get a better understanding of the gene function (supplemental Figure S2.5). Both, heterozygous *myh10*<sup>+/-</sup> and homozygous *myh10*<sup>-/-</sup> mutants are not lethal, survive until adulthood and become sexually mature. Surprisingly, no obvious anomalies were observed. Thus, to avoid maternal contribution, homozygous mutants were crossed to generated maternal-zygotic mutants. But unlike in morphants MZ*myh10*<sup>-/-</sup> zebrafish embryos developed normally and neither displayed an obvious ocular phenotype nor a muscle phenotype (supplemental Figure S2.6).



#### 4.2.4 Discussion – Project 2

We report on patients with a complex ophthalmic phenotype associating eyeball (coloboma) and periocular developmental anomalies (ptosis). Ocular coloboma is the result of a failure of the optic fissure closure occurring during the 5<sup>th</sup> to 7<sup>th</sup> week of human being's gestation affecting variably the iris and the retina with highly variable consequences ranging from asymptomatic to severe visual loss. Often associated to microphthalmia or other eye anomalies, this eye malformation can be due to environmental and genetic causes with high heterogeneity, with more than 30 genes identified to date involving mostly signaling pathways and transcription factors.<sup>273</sup> Syndromic forms are common with well characterized syndromes such as CHARGE syndrome (MIM # 214800), Branchio-Oculo-Facial syndrome (MIM # 113620). On another hand, congenital blepharoptosis is a common condition due to most often myogenic (or less frequent neurogenic) abnormal levator palpebrae superioris that can be related quite rarely to syndromic genetic conditions such as BPES (Blepharophimosis, Ptosis, and Epicanthus inversus Syndrome) (MIM # 110100), Saethre Chotzen syndrome (MIM # 101400) or Ohdo syndrome (MIM # 249620). Syndromes combining both eyeball and periorbital developmental anomalies have been very rarely described except for a few entities mostly associating coloboma and ptosis such as Noonan syndrome or BWCF and more recently in a new syndrome associating colobomatous-microphthalmia, ptosis, nephropathy and syndactyly due to biallelic variants in *FAT1* gene.<sup>283</sup>

Here, we report on a family with an autosomal dominant phenotype associating severe congenital ptosis and eyeball developmental anomalies. The clinical overlap with BWCF features was strong enough to consider the clinical presentation of the 3 patients as very closely related to this syndrome, although no neurological features are evidenced to date. Baraitser-Winter syndrome, first described in 1988, is indeed characterized by a facial gestalt (metopic ridging/trigonocephaly, bilateral ptosis and hypertelorism), intellectual disability and structural brain anomalies (agyria/pachygyria).<sup>277,284</sup> A common pathogenesis mechanism is underlying the BWCF clinical spectrum, with gain of function variants in two ubiquitous cytoplasmic actin-encoding genes namely *ACTG1* and *ACTB* that encode  $\gamma$  actin and  $\beta$  actin respectively.<sup>285</sup> Thus BWCF is now recognized as a syndrome with a wider phenotypic spectrum belonging to actinopathies<sup>277</sup> (see Table 16 for the summary of all phenotypic traits related to actinopathies). The patients described herein had mainly a cranio facial phenotype with bilateral ptosis and with eyeball malformations including colobomas (Figure 20). No intellectual deficiency was observed and magnetic resonance imaging did not reveal typical reported cerebral malformation previously described for BWCF. Ocular features of BWCF do include as a major feature ptosis and in 30 % of patients eye developmental anomalies such as coloboma affecting iris and/or retina and much more rarely microphthalmia or other eye malformations.

With the hypothesis of an actinopathy in the family, we focused the WES analysis first on genes involved in the cytoskeleton but no variants in the known BWCF genes or other genes involved in ocular developmental anomalies were detected. Interestingly, the three patients carried a heterozygous variant in the *MYH10* gene, encoding the non-muscle myosin heavy chain IIB (NMHC IIB) isoform, an actin based molecular motor. Moreover, MYH10 interacts with *KIF26B*, a gene in which a potential variant was reported in a patient with a renal coloboma syndrome.<sup>286</sup>

The 3 non-muscle myosin heavy chain II (NMII) genes, *MYH9*, *MYH10* and *MYH14*, encoding NMHC IIA, IIB, and IIC respectively, have overlapping and also unique properties as demonstrated by the related human phenotypes described to date. The actin motor proteins NMHC II are central in the control of cell adhesion, cell migration and tissue architecture.<sup>275,287,288</sup> They act as master regulators of cell morphology<sup>288</sup>, with a role in several essential cellular processes, including post-synaptic dendritic spine plasticity in neurons. Thus, NMHC II are at the forefront for pathogenesis studies as their activity is affected in a wide range of diseases such as neuronal disorders, cancer and cardiovascular diseases.<sup>289</sup> Pivotal roles of NMII during development have been investigated and highlight the impact of NMHC II isoforms during embryogenesis and organogenesis. The NMHC IIB myosin encoded by the *MYH10* gene is preferentially localized to the rear of the cell in migratory cells.<sup>290</sup> In mice, MYH10 is specifically enriched in neuronal tissue during mouse embryonic development and is critical for neuronal cell migration and for reorganization of the actin cytoskeleton.<sup>275,276</sup> Two *de novo* heterozygous variants of unknown significance in *MYH10* have been previously mentioned in patients presenting with a severe neurodevelopmental phenotype and brain anomalies but no further studies were performed leaving their pathogenic role uncertain.<sup>291,292</sup> Recent studies resulting from an ENU mice screen for modulators of lung development identified a homozygous missense *myh10* variant in the motor domain of the protein, associated to abnormal development of the alveolar epithelium leading to early death before birth<sup>293</sup>, suggesting that MYH10 plays a major role in developmental processes.

The actin cytoskeleton architecture is regulated by a large number of proteins that modulate filament assembly and disassembly.<sup>294</sup> It is also driven by the contractility and the filament disassembly mediated by myosin motor proteins.<sup>295,296</sup> The *MYH10* variant described here is localized in the tail domain (Figure 21). NMII myosins are characterized by a motor domain binding to F-actin, a neck domain composed of IQ motifs, and a tail domain of variable length and function that is required for assembly of the myosin II filament. A decrease in MYH10 protein level was observed in the presence of the variant, whereas the RNA level was not affected. This could be due to the Arg1471Pro missense mutation affecting the stability of the protein. This substitution could lead to a structural rearrangement due to the proline cyclic residue having a steric impact on the polypeptide chain<sup>297,298</sup> and a destabilizing effect on the coiled-coil domain in the mutated MYH10 protein. The

immunofluorescence studies on patient fibroblasts reveal a MYH10 mis-localization with less myosin filaments compared to control cells. The patient fibroblasts are also characterized by a lack of MYH10-actin filaments colocalization, a reduced of actin fiber length and abnormal structure of the cytoskeleton network (Figure 22 A-C and supplemental Figure S2.1). Spatiotemporal coordination of the Arp2/3 (actin-related protein 2/3) complex involved in actin polymerization and non-muscle myosin II contractile activity is important for neuronal growth cone actin dynamics.<sup>299</sup> Our data show that Arp3, one of the effector of this complex, was redistributed into the cytoplasm and no longer colocalized with actin filaments in patient fibroblasts, suggesting abnormal actin cytoskeleton dynamics (Figure 22 D). All these results are compatible with an actinopathy phenotype. Although the zebrafish morphants did not present a distinct coloboma phenotype, the development of the eye and the optic fissure closure were delayed by the down regulation of *myh10* (Figure 23A and B, supplemental Figure S2.2). Interestingly, *myh10* morphants additionally showed an impaired muscle structure (Figure 23 E-G). In zebrafish, *myh10* is expressed in the pre-myofibril, a precursor of the skeletal muscle.<sup>272</sup> The downregulation of *myh10* seems to be sufficient to affect skeletal muscle architecture. The fact, that *myh10*<sup>-/-</sup> CRISPR mutants did not display any obvious anomalies might be explained by genetic compensation and will be further discussed in the general discussion.

Overall, we show here that *MYH10* is involved in a human condition closely related to BWCFE actinopathy spectrum with the seldom reported association of eyeball developmental anomalies and congenital ptosis. MYH10 dysfunction leads to abnormal actin length and polymerization *in cellulo* and delayed development of the eye as well as a muscular phenotype in the zebrafish model. This report describing similar phenotypes due to mutation in *MYH10* myosin and *ACTB* or *ACTG1* actin genes should raise awareness of analyzing effectors of the actomyosin network to detect additional patients and to expand the clinical and genomic spectrum related to BWCFE.

#### 4.2.5 Supplementary Data -Project 2 (S2)

##### **Authors contributions**

S.S., E.S., S.K. and H.D. gathered data from patients and performed clinical investigations; J.M. and V.G. gathered sequencing data and performed analyses; **A.KH.**, S.B., C.E, C.S and S.F. performed cell biology experiments and data analyses; **A.KH.**, C.E. and U.S. designed and performed the zebrafish experiments and data analyses; **A. KH.**, S.S., S.B, S.F., J.M. and H.D. analyzed the data and wrote the paper; C.E., C.S. and U.S. contributed to manuscript writing; S.F., U.S. and H.D provided direction for the project, conceived and designed the experiments.

##### **Supplemental Methods**

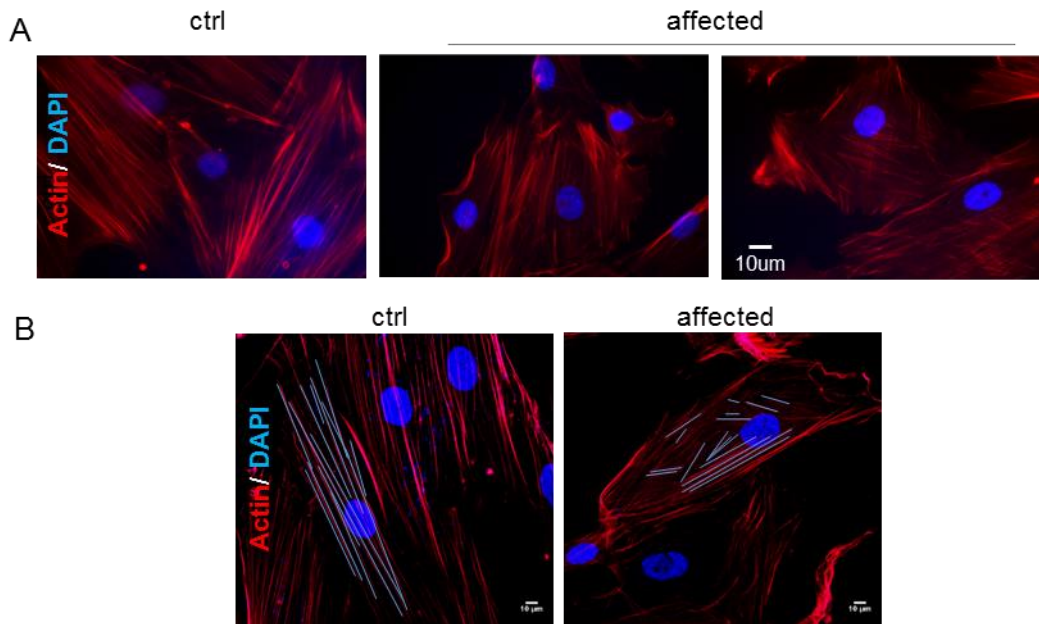
###### **Method S2.1 Candidate gene approach: *in silico* prediction of potential genes implicated in the BWCF syndrome**

BWCF syndrome is due to pathogenic variants on either *ACTB* or *ACTG1* genes.<sup>285</sup> We listed the connected human genes to *ACTB* or *ACTG1* provided by the FunCoup 4.0 human database<sup>271</sup>, applying a confidence threshold of 0.8 (= minimum experimental support required for a link in FunCoup) and a single level of interaction.

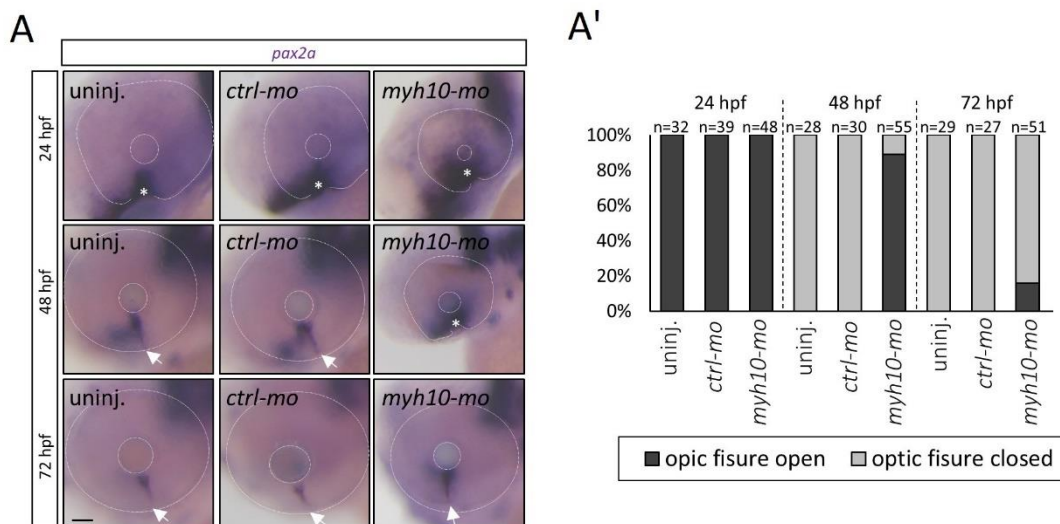
###### **Method S2.2 Morpholino efficiency test**

To verify the morpholino efficiency, we amplified and cloned a 392 bp fragment covering the 5'UTR and exon 1 of *myh10* cDNA (*myh10*<sub>5'UTR-ex1</sub>) into pCS2+gfp in such a way that GFP was in frame with the *myh10* fragment. RNA was synthesized using the mMESSAGING mMACHINE SP6 Kit (AM1340, Ambion) according to the manufacturer's instructions.

Supplemental Figures

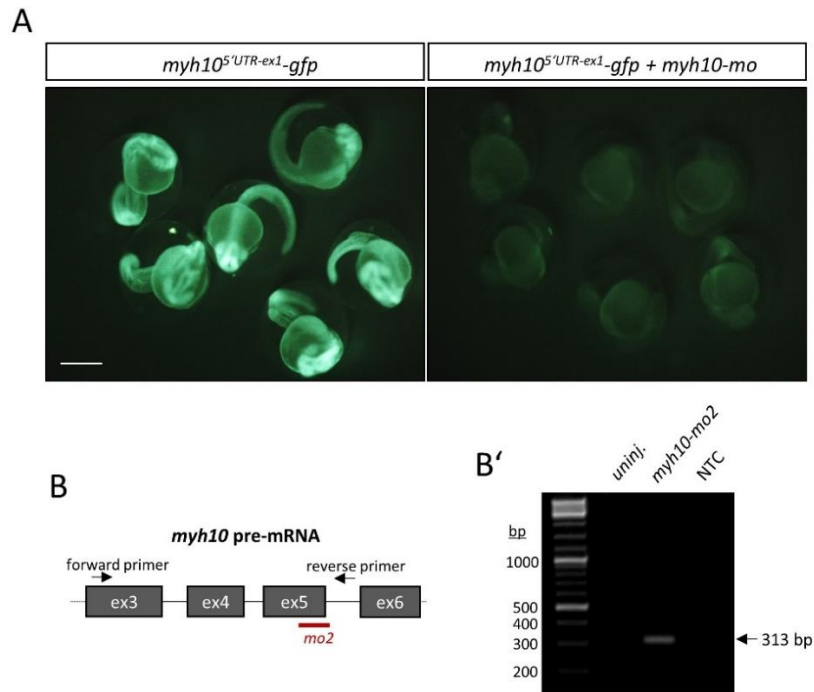


**Figure S2.1. Actin filaments are shorter in patient’s skin fibroblasts.** Cells were grown on coverslips, fixed, actin stained with Rhodamine-phalloidin and nuclei with DAPI. **(A)** actin distribution in control (ctrl) and affected skin fibroblasts was observed on a fluorescence microscope. **(B)** 15 to 20 filaments of five cells in five independent experiments were measured using ImageJ, mean length determined and statistical significance determined using a T-Test. \*:  $p < 0,005$



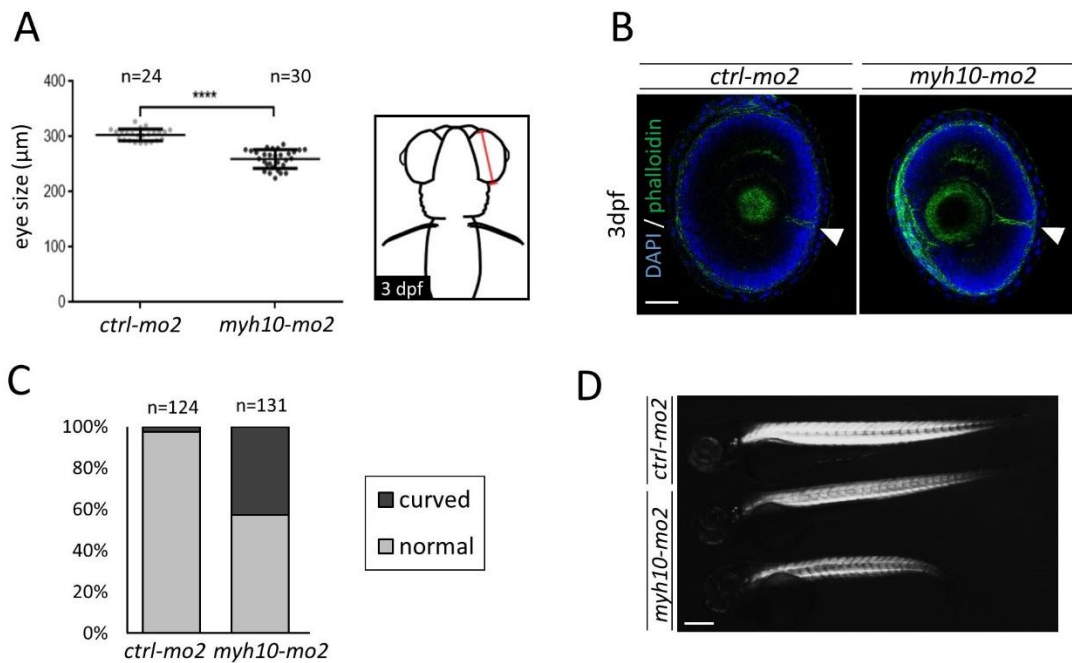
**Figure S2.2. Optic fissure closure is delayed in *myh10* morphants.**

**(A)** The diameter of the eye of 3-day-old *myh10* morphants are smaller compared to the eyes of control injected embryos. Statistical significance was determined using the Mann-Whitney Test, ns = non-significant. \*  $p < 0.05$ . \*\*  $p < 0.01$ . \*\*\*  $p < 0.001$ . \*\*\*\*  $p < 0.0001$ . **(B)** 3-day-old embryos were immunostained with phalloidin (in green) and DAPI (in blue) to visualize the optic fissure closure. The optic fissure of both, uninjected and *myh10-mo* injected embryos, was closed.

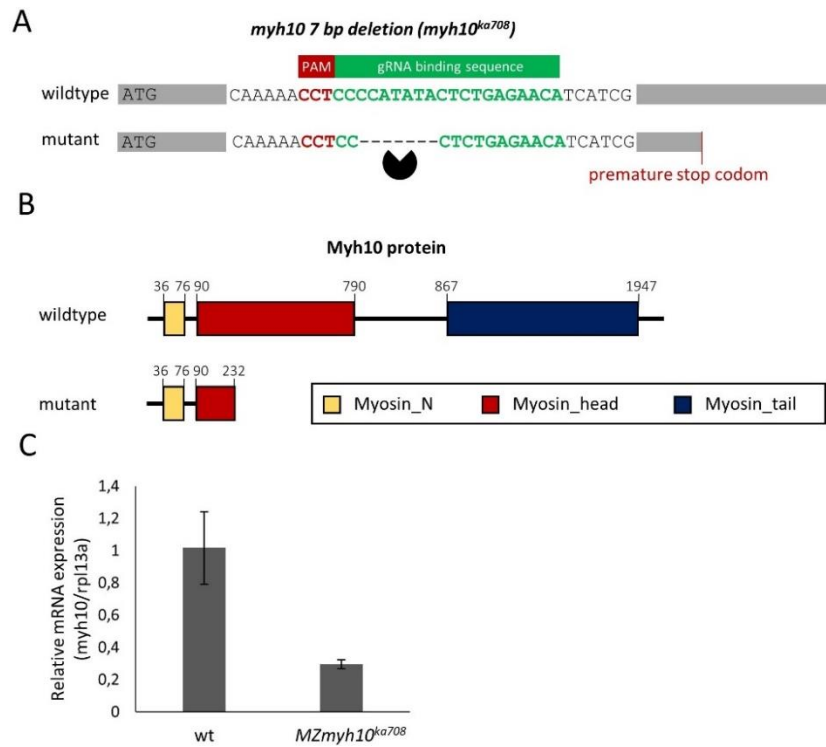


**Figure S2.3 Morpholino efficiency test of *myh10*-morphants.**

(A) The efficiency of the start-site morpholino (*myh10-mo1*) was validated by co-injecting a *myh10* N-terminal sequence fused to GFP. Mo1 efficiently binds to *myh10*<sup>5'UTR-ex1-gfp</sup> and prevents the expression of the construct. Scale bar 500  $\mu$ m. (B+B') To examine the efficiency of the splice-site *myh10*-morpholino (*mo2*), RNA was extracted from uninjected and *mo2*-injected embryos, a cDNA was synthesized and a RT-PCR was performed showing that the *myh10* pre-mRNA fails to splice the exon5-intron5-6 boundary in *myh10-mo2* injected embryos. NTC: no template control.

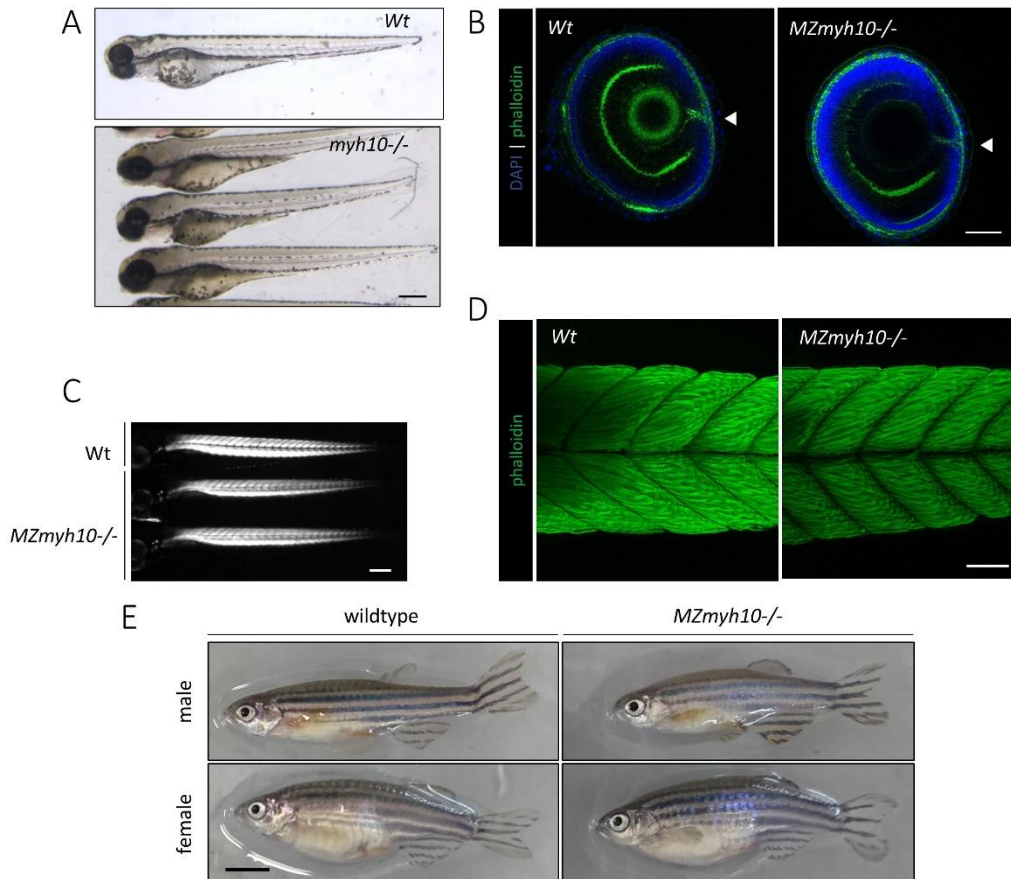


**Figure S2.4. A second morpholino (*myh10-mo2*) confirmed the eye and muscle phenotype.** A. 3-day-old *myh10-mo2* morphants displayed smaller eyes compared to control injected embryos. Statistical significance was determined using the Mann-Whitney Test, ns = non-significant. \*  $p < 0.05$ . \*\*  $p < 0.01$ . \*\*\*  $p < 0.001$ . \*\*\*\*  $p < 0.0001$ . B. Alike in *myh10-mo* morphants, the optic fissure of *myh10-mo2* morphants was closed after 3 dpf. Actin filaments were stained with phalloidin toxin (green) and cell nuclei with DAPI (blue). C. Quantification of embryos with a curved body axis. D. Alike in *myh10-mo* morphants, the muscle development of 3 dpf old embryos injected with *myh10-mo2* was impaired. Muscle integrity was assessed by birefringence. Scale bar 50  $\mu$ m.



**Figure S2.5 Generation of a *myh10* KO mutant using CRISPR/Cas9.**

**(A)** The guide RNA binds a sequence within exon 3 of the *myh10* gene and results in a deletion of seven base pairs resulting in a frameshift and a premature stop codon. **(B)** Schematic representation of the Myh10 protein structure of wildtype and *myh10*<sup>-/-</sup> mutants. **(C)** Real-time qPCR analyses of the expression of *MYH10*, normalized to rpl13a in wildtype and *myh10*<sup>-/-</sup> mutants. Error bars represent standard deviation of technical triplicates. n= 50 embryos (3 dpf)/replicate.



**Figure S2.6 *myh10*<sup>-/-</sup> mutants do not present any obvious phenotype.**

**(A)** Representative image of 3-days-old wildtype embryos and *myh10*<sup>-/-</sup> mutants. Scale bar 250  $\mu$ m. **(B)** 3-day-old embryos were immunostained with phalloidin (in green) and DAPI (in blue) to visualize the optic fissure closure. The optic fissure of both, wildtype and *myh10*<sup>-/-</sup> mutants, was closed. **(C)** A normal birefringence was observed in wildtype and mutant embryos indicating a normal muscle organization. Scale bar 50  $\mu$ m. **(D)** Immunostaining, using antibodies against myosin phalloidin toxin (in green) confirmed that the muscle organization is not affected in *myh10*<sup>-/-</sup> mutants. Scale bar 50  $\mu$ m. **(E)** Alike mutant embryos, *MZmyh10*<sup>-/-</sup> adults (12 mpf) did not display obvious eye or body malformations. Scale bar 500  $\mu$ m.







**Table S2.2 Summary of the whole exome sequencing results. No potentially pathogenic variant was identified in ACTB and ACTG1 genes.**

	III-1			III-2			II-2		
	SNV	indel	SV	SNV	indel	SV	SNV	indel	SV
Total number of variants	87894	12847	13	88879	13435	12	90465	13177	13
After exclusion of variants with an allele frequency > 1 % (gnomAD, 1000G, internal exome database, DGV)	7803	1118	7	7773	1138	4	8310	1144	7
After exclusion of SNV/indel found in the homozygous state in gnomAD and in our internal exome database	2944	449	-	2825	413	-	3254	443	-
SNV: Single nucleotide variation, indel: gain or loss of up to 50 nucleotides at a single locus, SV: Structural Variation. Exclusion of SV with a DGV frequency > 1 % is done only with studies of more than 1000 individuals.									

# Project 3

Based on:

Article was submitted to the journal EMBO Molecular Medicine

## Proteasome subunit PSMC3 variants cause neurosensory syndrome combining deafness and cataract due to proteotoxic stress

*Ariane Kröll-Hermi<sup>1,2\*</sup>, Frédéric Ebstein<sup>3\*</sup>, Corinne Stoetzel<sup>1\*</sup>, Véronique Geoffroy<sup>1\*</sup>, Elise Schaefer<sup>1,4</sup>, Sophie Scheidecker<sup>1,5</sup>, Séverine Bär<sup>6</sup>, Masanari Takamiya<sup>2</sup>, Koichi Kawakami<sup>7</sup>, Barbara A. Zieba<sup>3</sup>, Fouzia Studer<sup>8</sup>, Valerie Pelletier<sup>4,8</sup>, Claude Speeg-Schatz<sup>9</sup>, Vincent Laugel<sup>1,10</sup>, Dan Lipsker<sup>11</sup>, Florian Sandron<sup>12</sup>, Steven McGinn<sup>12</sup>, Anne Boland<sup>12</sup>, Jean-François Deleuze<sup>12,13</sup>, Lauriane Kuhn<sup>14</sup>, Johana Chicher<sup>14</sup>, Philippe Hammann<sup>14</sup>, Sylvie Friant<sup>6</sup>, Christelle Etard<sup>2</sup>, Elke Krüger<sup>3\*</sup>, Jean Muller<sup>1,5\*</sup>, Uwe Strähle<sup>2\*</sup>, Hélène Dollfus<sup>1,4,8\*</sup>*

\* equal contributors

1. Laboratoire de Génétique Médicale, INSERM, UMRS\_1112, Institut de Génétique Médicale d'Alsace (IGMA), Université de Strasbourg Faculté de médecine de Strasbourg, 11 rue Humann, Strasbourg, France.
2. Karlsruhe Institute of Technology (KIT), Institute of Toxicology and Genetics (ITG), Eggenstein-Leopoldshafen, Germany.
3. Institut für Medizinische Biochemie und Molekularbiologie (IMBM), Universitätsmedizin Greifswald, Ferdinand-Sauerbruch-Straße, Klinikum DZ 7, 14175 Greifswald, Germany
4. Service de Génétique Médicale, Hôpitaux Universitaires de Strasbourg, Strasbourg, France.
5. Laboratoires de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France.
6. Laboratoire de Génétique Moléculaire, Génomique, Microbiologie (GMGM), UMR7156, Centre National de Recherche Scientifique (CNRS), Université de Strasbourg, Strasbourg, France.
7. Laboratory of Molecular and Developmental Biology, National Institute of Genetics, and Department of Genetics. SOKENDAI (The Graduate University for Advanced Studies), Mishima, Japan
8. Centre de Référence pour les affections rares en génétique ophtalmologique, CARGO, Filière SENSGENE, Hôpitaux Universitaires de Strasbourg, Strasbourg, France.
9. Department of Ophthalmology, Hôpitaux universitaires de Strasbourg, Strasbourg, France
10. Service de Pédiatrie, Hôpitaux Universitaires de Strasbourg, 1 avenue Molière, Strasbourg, France.
11. Faculté de Médecine, Université de Strasbourg et Clinique Dermatologique, Hôpitaux Universitaires, Strasbourg, France.
12. Centre National de Recherche en Génomique Humaine (CNRGH), Institut de Biologie François Jacob, CEA, Université Paris-Saclay, F-91057, Evry, France.
13. Centre d'études du polymorphisme humain-Fondation Jean Dausset, Paris, France.
14. Institut de Biologie Moléculaire et Cellulaire (IBMC), Plateforme Protéomique Strasbourg—Esplanade, CNRS FRC1589, 67084 Strasbourg, France

## 4.3 Project 3

Parts of this chapter (4.3.2-4.3.5) were reproduced with permission from the article: **“Proteasome subunit PSMC3 variants cause neurosensory syndrome combining deafness and cataract due to proteotoxic stress”** that has been submitted to the journal **“EMBO Molecular Medicine”**. The synopsis will give a brief summary of the project and points out my personal contribution in this work.

### 4.3.1 Synopsis

Early-onset cataract and congenital deafness is a very rare combination of symptoms, that has been observed only in very few syndromes, e.g. the Aymé-Grip syndrome and the Wolfram syndrome.<sup>300,301</sup> In project 3, I report a large inbred family of Turkish origin with three children affected by early-onset cataract, congenital deafness and autism. To identify the pathogenic gene mutation, WES in three patients was performed but failed to reveal a candidate gene that might cause the observed combination of symptoms. Therefore, WGS was carried out in all affected and two healthy individuals, which led to the identification of a homozygous deep intronic mutation in the gene *PSMC3* (c.1127+337A>G), predicted to create a new splice donor site.

The *PSMC3* gene encodes the 26S protease regulatory subunit 6A (Rtp5) and is one of the six known 26 proteasome AAA-ATPase subunits forming the ATPase Ring (Rpt1-6). As part of the ubiquitin proteasome system (UPS), *PSMC3* is involved in the regulation of intracellular protein levels and protein quality control and thus plays an important role in cell homeostasis.<sup>43</sup>

To date, dysfunctions of the UPS have been especially observed in neurodegenerative diseases, such as Alzheimer disease, Huntington disease (MIM: 143100), Parkinson disease (MIM: 168600) or amyotrophic lateral sclerosis (MIM:105400).

In order to investigate the effect of the deep intronic *PSMC3* mutation found in all three affected individuals on the mRNA expression, an RT-PCR was conducted and confirmed the presence of the wildtype mRNA as well as an alternative spliced form in patient's cells. Alternative splicing is caused by the deep intronic mutation that leads to the inclusion of a cryptic exon of the size of 114 bp and thus results in a premature stop codon. Moreover, a quantitative real-time PCR was performed and revealed a downregulation of the *PSMC3* wildtype mRNA in patients' cells as well as the presence of an additional truncated form. The protein level and the localisation of *PSMC3*, however, seemed to be not affected by the deep intronic mutation. The truncated protein seems to be unstable, as it was not detectable by Western Blot.

Considering the importance of *PSMC3* for the degradation of ubiquitinated proteins, the impact of the deep intronic mutation on the UPS function was investigated. Interestingly, in normal conditions, fibroblasts derived from patients carrying the homozygous *PSMC3* mutation exhibited an increased amount of proteasome complexes. The proteasome function seemed to be not affected. Under stress

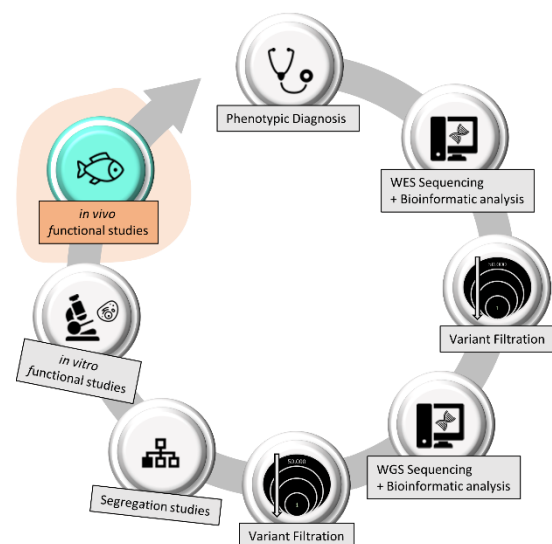
conditions, however, patient's fibroblasts exhibit an exhausted TCF11/Nrf1 processing pathway which prevents them to upregulate proteasome subunits and thus fail to preserve protein homeostasis as a result of proteasome deficiency.

Moreover, the zebrafish as a known animal model for human eye diseases and deafness, was used to study the link between cataract/deafness and *PSMC3* deficiency. The *psmc3* zebrafish gene has a high sequence identity of 83 % to the human orthologue. To generate a *psmc3* knockdown model, I used a morpholino antisense oligonucleotide targeting a splice-site in *psmc3*. After 4 dpf, the development of the zebrafish lens and ear were assessed in living embryos. In *psmc3* morphants, a strong cataract phenotype, as well as abnormalities of semicircular canal fusion in the zebrafish's ear, were observed. In contrast, embryos injected with a control morpholino (with a 5 base pair exchange) developed normally. To evaluate autistic traits, I performed a whole mount in situ hybridization using brain markers, but I was unable to observe an obvious difference between morphants and a control group. To confirm these results, I used the CRISPR/Cas9 technique to edit the zebrafish genome. The F0-Generation is often genetically mosaic, however, in the case of highly efficient guide RNAs, it has been shown that F0s, also called crispants, can successfully recapitulate mutant phenotypes.<sup>8-10</sup> Also, in this project, crispants recapitulated the eye and ear phenotype observed in *psmc3* morphants. A rescue experiment, confirmed the specificity of morpholino and guide RNAs, as co-injection of *psmc3* messenger RNA could rescue the eye and ear phenotype in morphants and crispants.

Overall, we conclude that PSMC3 deficiency is associated with the development of cataract and deafness and suggest an important role of PSMC3 during the lens and ear development of humans and zebrafish.

### Personal contribution

For project 3, I performed all zebrafish experiments, partly with the support of Masanari Takamiya and Christelle Etard and was implicated in the scientific discussion, and the writing of the publication (Figure 24). For a detailed description of the authors contributions see the supplemental Information of Project 3 (S3).



**Figure 24. Personal contribution to project 3.** Icons were created with icons8.com.

#### 4.3.2 Introduction - Project 3

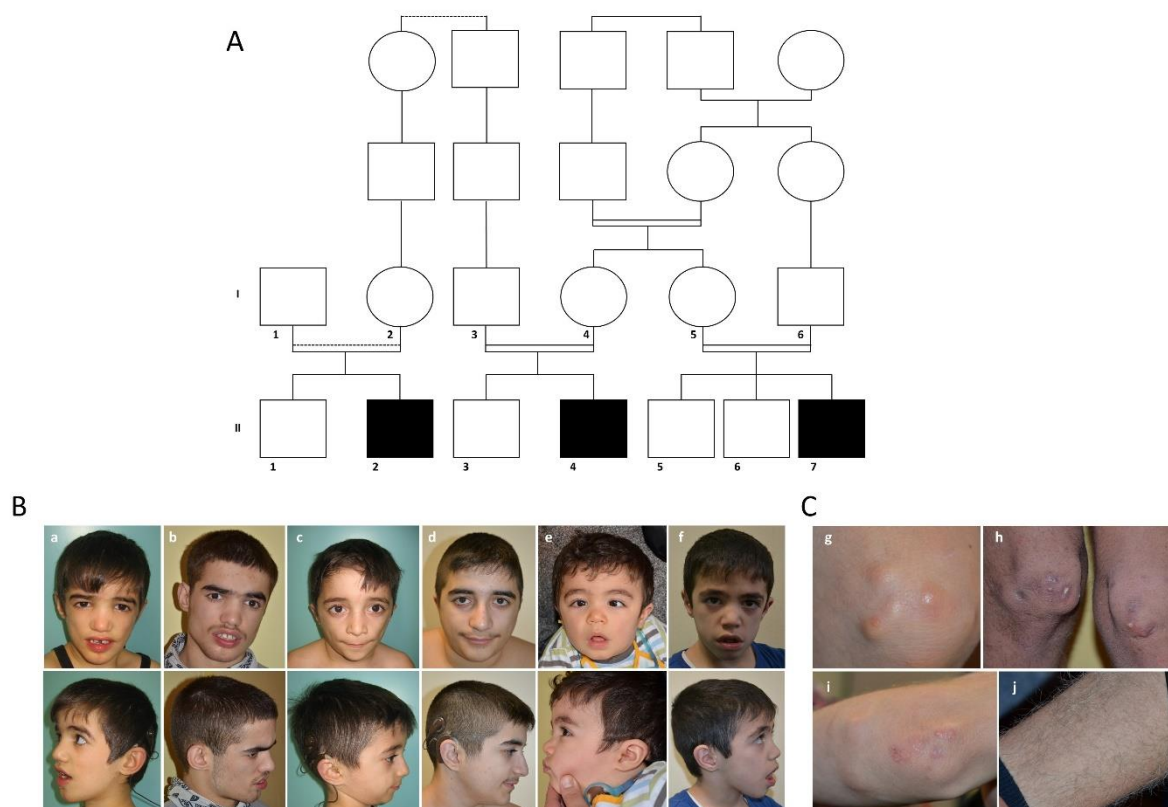
Early onset deafness is one of the most common causes of developmental disorder in children (prevalence rate of 2-4/1000 infants) and identically early-onset cataract is the most important cause of paediatric visual impairment worldwide (prevalence form 2-13.6/10000 according to regions) accounting for 10 % of the causes of childhood blindness. Each condition can be attributed to environmental causes (intrauterine infections, inflammation, trauma or metabolic diseases) or to genetic causes with a well-recognized very high level of genetic heterogeneity with 59 known genes causing early-onset cataracts and 196 genes known to cause severe deafness<sup>302,303</sup>. Patients presenting both entities simultaneously, early onset severe deafness and congenital cataracts, are thought to be mainly due to teratogenic exposure during pregnancy especially infections and are, nowadays, considered to be very rare. Indeed, only very few genetic inherited entities associating both congenital cataracts and deafness have been reported so far. The Aymé-Gripp syndrome (cataract, deafness, intellectual disability, seizures and Down syndrome like facies) has been recently linked to *de novo* pathogenic variants in the *MAF* gene a leucine zipper-containing transcription factor of the AP1 superfamily.<sup>300</sup> In addition, dominant pathogenic variants in *WFS1* (recessive loss of function variants are responsible for Wolfram syndrome) have been described in children with congenital cataracts and congenital deafness presenting in the context of neonatal/infancy -onset diabetes<sup>301</sup>.

Herein, using whole-genome sequencing, we describe a novel homozygous non-coding pathogenic variant in *PSMC3* associated with severe congenital deafness and early onset cataracts and various neurological features in 3 patients from a very large consanguineous family. *PSMC3* encodes the 26S regulatory subunit 6A also known as the 26S proteasome AAA-ATPase subunit (Rpt5) of the 19S proteasome complex responsible for recognition, unfolding and translocation of substrates into the 20S proteolytic cavity of the proteasome<sup>304</sup>. The proteasome is a multiprotein complex involved in the ATP-dependent degradation of ubiquitinated proteins to maintain cellular protein homeostasis and to control the abundance of many regulatory molecules. We suggest that biallelic loss of *PSMC3* causes a novel autosomal recessive syndrome with varying degrees of neurosensorial dysfunctions including the combination of cataract and deafness. Functional analysis of patient's cells revealed that although normal amount of proteasome proteins can be observed in steady-state conditions, the cells are unable to adapt to proteotoxic stress. The use of zebrafish morpholinos and CRIPSR-Cas9 assays confirmed the same combination of sensory phenotypes upon inactivating *PSMC3* expression.

## 4.3.3 Results - Project 3

**Patient phenotypes**

Three patients with a novel syndromic neuro-sensory-cutaneous presentation consulted independently to our clinical centre over a period of 15 years. Careful analysis revealed that they originated from the same small village (Amarat) in the Kayseri region of Turkey and belong to the same large extended consanguineous family (Figure 25A).



**Figure 25. Family pedigree with three affected individuals.**

**(A)** Family pedigree. **(B)** Face (up) and profile (down) photographs for patients II.4 (a: 8 yo, b: 16 yo), II.2 (c: 6 yo, d: 14 yo) and II.7 (e: 1 yo, f: 7 yo) over time. **(C)** Sub cutaneous calcifications found only on knees (g: 9 yo and h: 16 yo) on elbows (i: 9 yo) of patient II.4. White hairs were present only on the legs of the 3 patients as illustrated in patient II.4 (j: 16 yo).

The proband is a male individual (II.4) diagnosed at the age of 7 months with severe perceptive deafness and subsequently benefited from a cochlear implantation. He was referred at the age of 2 years old to our center because of visual impairment due to bilateral cataracts for which he underwent bilateral lensectomies. With years, he developed severe developmental delay and severe intellectual deficiency (no words, limited comprehension). Several facial features were noted (Table 17 and Figure 25B). In addition, autistic features and peripheral polyneuropathy of lower limbs were revealed at the age of 2.5 years old (Table 17). A full metabolic exploration was normal. At the age of 5 he developed sub cutaneous deposits at the level of the knees and elbows (Figure 25C). At the age of ten he



developed white hair at the level of the two legs as opposed to the dark pigmented hair on the rest of the body.

The 2 other patients (II.2 and II.7) were referred at the age of one year old and share the same severe congenital perceptive deafness (for which they also benefited from a cochlear implantation) and visually impairment due to bilateral obstructive cataracts (for which they also had bilateral lensectomies). Patient II.2 did not present with autistic features but moderate developmental delay. Patient II.7 did not have peripheral polyneuropathy of lower limbs compared to the other patients.

**Table 17. Clinical description of the patients with PSMC3 pathogenic variants.** M: moderate, S: severe

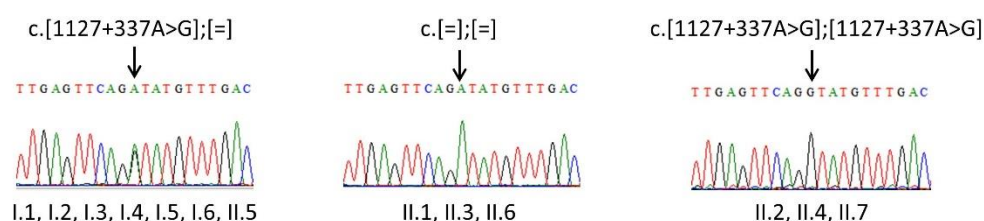
<b>Individual</b>	<b>II.4</b>	<b>II.2</b>	<b>II.7</b>
<b>Birth date</b>	24/12/2003	02/02/2005	15/03/2012
<b>Neurosensory features</b>			
Congenital cataract	+	+	+
Strabismus	+	-	+
Congenital deafness	+	+	+
<b>Facial dysmorphism features</b>			
Round Ears with no lob	+	-	-
Synophris	+	-	-
Short philtrum	+	+	-
Malar region: retrusion	+	+	+
Prominent supraciliary arches	+	+	+
Sunken cheeks	+	+	+
Preauricular fibrochondroma	-	-	+
Thin upper lip	+	+	-
<b>Neurologic features</b>			
Developmental delay	S	M	S
Autistic features	+	-	+
Peripheral polyneuropathy of lower limbs	+	+	-
<b>cutaneous features</b>			
Calcifications of elbows and knees	+	+	+
Depigmented hairs of lower limbs	+	+	-

### Identification of a rare deep intronic variation in *PSMC3*

During the years of follow-up each patient was explored for known deafness and cataract genes by Sanger sequencing (in particular for *GJB2*, one of the patient being an heterozygous carrier of the c.30delG well known pathogenic variant) but also using larger assays such as Whole-Exome Sequencing (WES) with a specific focus on known deafness and cataract genes (supplemental Table S3.1) and standard chromosomal explorations (karyotype and chromosomal microarray analysis) but all were negative (see supplementary methods S3.1). Considering that affected individuals may harbor pathogenic variants in a region not covered by the WES (i.e. intronic, intergenic...) or not well detected (i.e. structural variations)<sup>186</sup>, we applied whole-genome sequencing (WGS) to the three affected individuals (II.2, II.4 and II.7) and two healthy individuals (II.1 and II.3). Given the known consanguinity in the family, our analysis was focused on homozygous variations and more specifically within

homozygous regions defined by the SNP arrays (supplemental Figure S3.1). In addition to the classical filtering strategy including functional criteria, frequency in population-based databases and cosegregation analysis, we defined a list of 4846 potentially interacting genes with the already known human cataract (59) and deafness genes (196) (supplemental Figure S3.2). This strategy allowed us to identify from the ~5,000,000 variations per WGS, 6 variations out of which a unique homozygous variant in the intron 10 of the *PSMC3* gene (c.1127+337A>G, p.?) remained of interest (supplemental Table S3.2 and supplementary methods S3.2). This variant was not present in any variation database (e.g. gnomAD) and is predicted to create a new donor splice site. Interestingly among others *PSMC3* was shown to interact (supplemental Table S3.3 and S3.4, supplemental Figure S3.3) with *CHMP4B* (MIM 610897), *ACTG1* (MIM 102560) and *GJB6* (MIM 604418) involved in cataract and deafness.

The variant confirmation and the cosegregation analysis with the phenotype in the family member were performed by Sanger sequencing after PCR amplification (Figure 26).



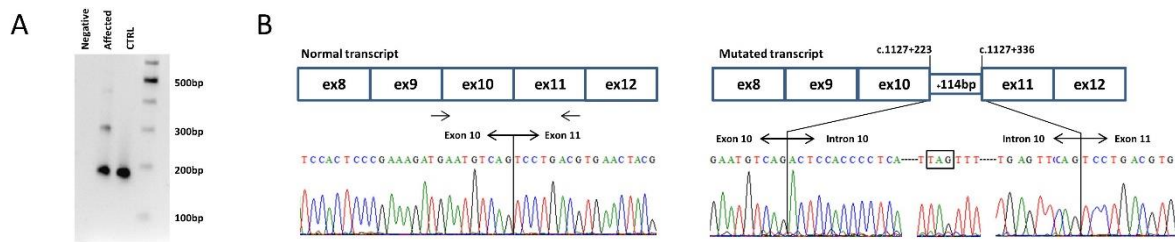
**Figure 26. Mutation segregation analysis of *PSMC3* intron 10 mutation.**

Electropherogram of a part of intron 10 of *PSMC3* encompassing the identified mutation c.[1127+337A>G];[1127+337A>G], p.[(Ser376Arg15\*);(Ser376Arg15\*)] in the affected individuals, their unaffected parents and siblings. The mutation was found at the homozygous state in the affected individuals (II.2, II.4, II.7), at the heterozygous state in the parents (I.1, I.2, I.3, I.4, I.5, I.6) and was either at the heterozygous state (II.5) or absent in the unaffected siblings (II.1, II.3, II.6).

### Effect of the variation on *PSMC3* expression and localization

In order to prove the effect of this deep intronic variation, we investigated the expression of the gene in the patient's fibroblasts. The suspected new donor site could be associated to multiple acceptor sites within the intron 10 (supplemental Figure S3.4). RNA analysis revealed an additional band specific to the affected patient that was further explored by Sanger sequencing (Figure 27A). The consequence of this variation is the inclusion of a cryptic 114 bp exon during the splicing process based on the intronic sequence (r.1127\_1127+1insACTCCACCCCTCATCTGAAGGCACAGAGGCTGGAGGCACTTAGTTT CCTGGCCTCACACCTCAGCCCATTAACACACGCCAGGAATGGCCGGGACCAGATGGACTTGAGTTTCAG) (Figure 27B) that is predicted to add 15aa (LeuHisProSerSerGluGlyThrGluAlaGlyGlyThr) at position 376 followed by a stop (p.(Ser376Argfs15\*)). Analysis at the RNA level showed a significantly reduced level of *PSMC3* as well as the presence of an additional truncated form (supplemental Figure S3.5). However, no difference in expression or localization of the *PSMC3* protein could be detected between the control

and the patient cells under normal condition indicating that the truncated form is probably not stable (supplemental Figure S3.6).

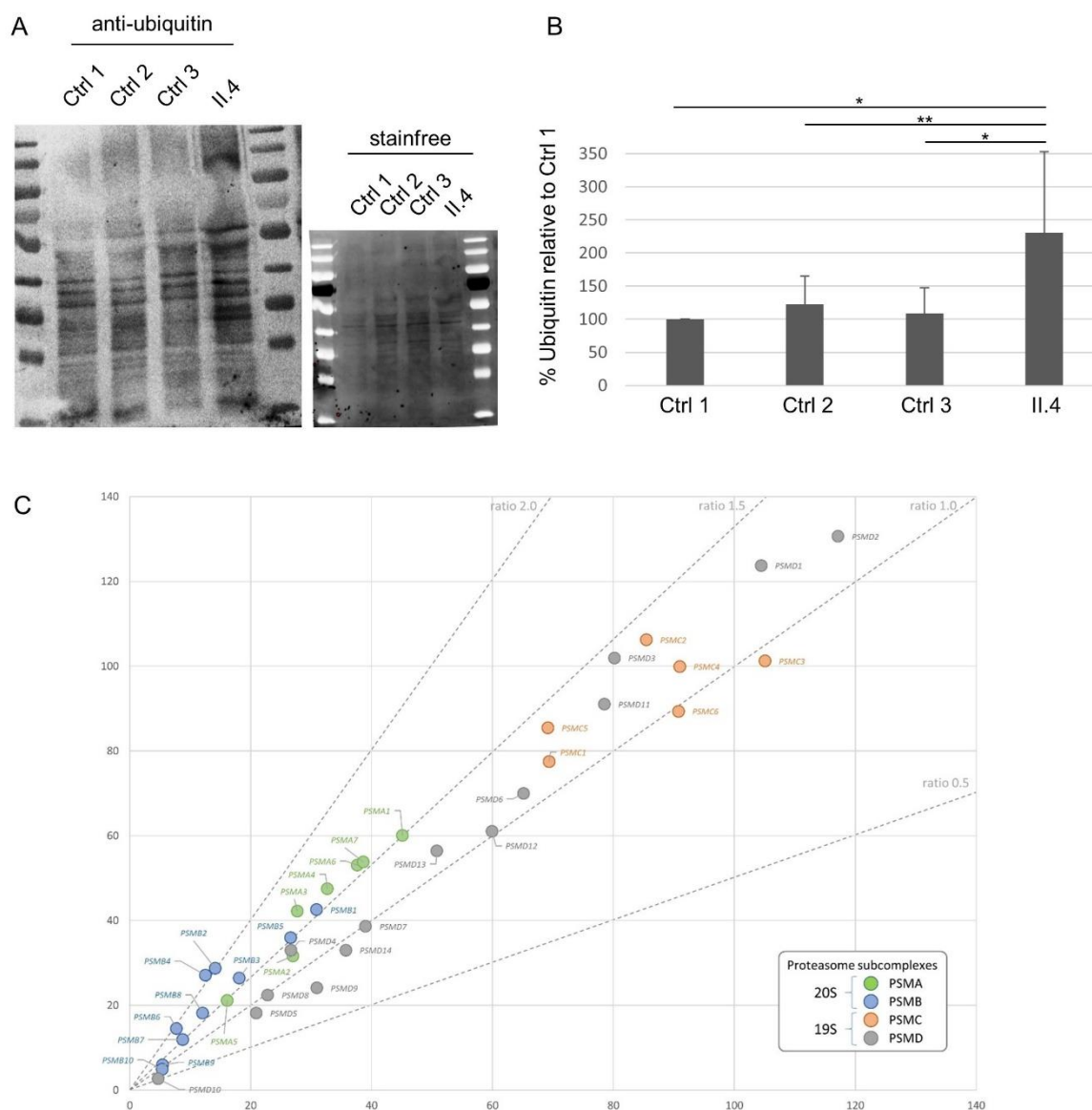


**Figure 27. cDNA analysis.**

**(A)** Amplification of the cDNA fragment between exons 9-10 and 11 of PSMC3 showing the abnormally spliced RNA fragment. One band at 180bp representing the normal allele is seen for the control and two bands for the individual II.4 (pathologic allele at 300bp). **(B)** Schematic representation for the incorporation of the 114bp intronic sequence resulting from the c.1127+337A>G deep intronic mutation on the mRNA. Sanger sequencing of the fragment between exons 9-10 and 11 of the PSMC3 cDNA obtained from patient II.4 fibroblasts' RNA, showing the insertion of the 114bp cryptic exon. As a comparison, the schematic representation and sequence from a control individual are shown below.

### Functional effect of the PSMC3 variant to the proteasome function and assembly

Given the role of PSMC3 in protein degradation, we determined the intracellular level of ubiquitinated proteins in patient cells compared to controls. Our results show an increased level of ubiquitinated proteins in patient cells (Figure 28A and B), suggesting that the proteasomal proteolysis is less efficient. Having shown a possible effect on proteasome function, we next investigated how the variant could affect the proteasome assembly and dynamics. First, in standard condition, PSMC3 protein and its partners were immunoprecipitated from either controls' or patient's fibroblasts and revealed by mass spectrometry (Figure 28C). The PSMC3 protein was detected in the input control and patient indicating that the variant does not affect the protein stability of the remaining wild type allele confirming the western-blot analysis. Looking at the interacting partners, one can notice that each proteasome subunit could be detected revealing no apparent defect in the general organization of the proteasome. However, protein abundance of each proteasome subcomplex estimated from the number of mass spectrometry spectra observed between the controls and the patient (supplemental Table S3.5) shows a general increase of the proteasome subunits (approximately 20 %). Looking more specifically at each sub complex, the increase is mainly due to the core particle including the alpha and beta rings with 1.5-fold for all PSMA proteins and most PSMB protein and even a 2.0-fold increase for PSMB2/4/6 while PSMC and PSMD remain at the same ratio.

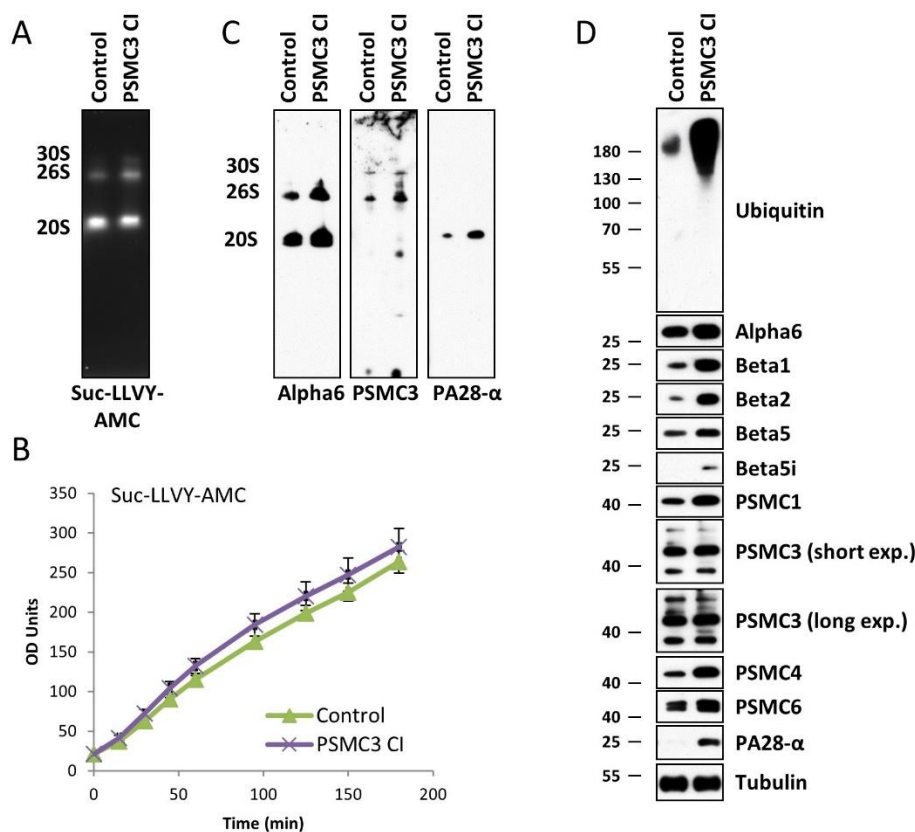


**Figure 28. Effect of the deep intronic PSMC3 variant on the proteasome function.**

**(A)** Anti-ubiquitin western blot in control and patient fibroblasts and total amount of proteins loaded (stainfree) showing increased ubiquitination in patient cells (lane 4). **(B)** Histogram showing the quantification of ubiquitin with western blot assays. The data shown correspond to the sum of all bands detected by the anti-ubiquitin antibody expressed as a percentage of the amount of ubiquitin in “Control 1” cells. Results are the mean of ten independent experiments. \*:  $p < 0.01$ , \*\*:  $p < 0.05$ . **(C)** Mass spectrometry results from the co immunoprecipitation with PSMC3 are displayed as the normalized total number of spectra count of each protein computed as the mean from 3 controls (x axis) vs the mean of patient II.4 triplicate. Proteasome subcomplexes are colored according to the displayed legend and standard ratio lines are drawn.

To further characterize the consequences of the *PSMC3* variant on the functionality of the ubiquitin-proteasome system (UPS), cell lysates derived from control and patient primary fibroblasts were analyzed by non-denaturing PAGE with proteasome bands being visualized by their ability to hydrolyse the fluorogenic peptide. As shown in Figure 29A, gel overlay assay for peptidase activity revealed two

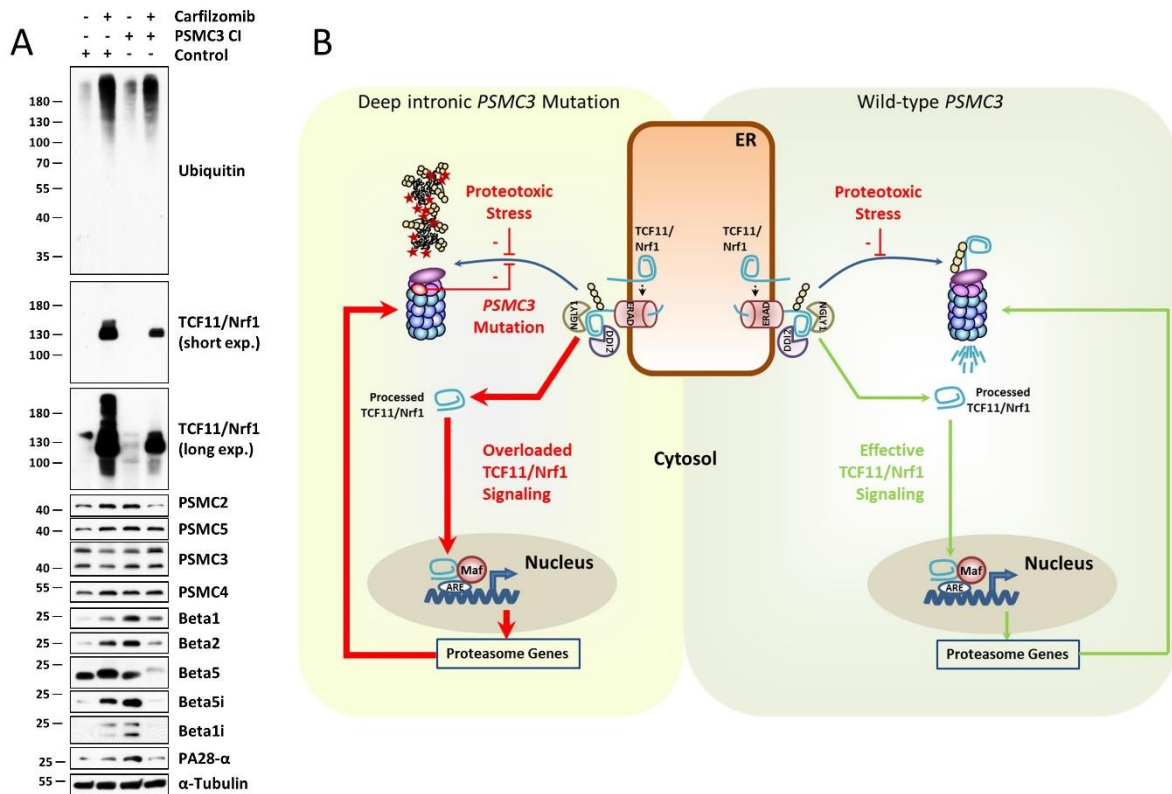
strongly stained bands corresponding to the positions of the 20S and 26S proteasome complexes, respectively. However, no discernible differences could be detected in the chymotrypsin-like activity of both 20S and 26S complexes between control and patient cells, suggesting that peptide hydrolysis in the 20S proteolytic core is not substantially impaired by the *PSMC3* pathogenic variant. The notion that proteasome activity does not vary between these two samples was further confirmed by monitoring the degradation rate of the Suc-LLVY peptide directly in whole-cell extracts from control and patient fibroblasts over a 180-min period that was almost identical in both samples (Figure 29B). In order to characterize the proteasome populations in cells carrying the deep intronic *PSMC3* homozygous variant, samples separated by non-denaturing PAGE were subsequently analyzed by western blotting. As expected, using an antibody against the proteasome subunit Alpha6, two major bands were observed in the 20S and 26S regions (Figure 29C). Interestingly, the signal intensity for both proteasome complexes was significantly stronger in patient's fibroblasts, suggesting that the amount of intracellular proteasome pools in these cells was higher than those of control fibroblasts. Western blotting against the *PSMC3* subunit revealed two bands in the 26S proteasome area and corresponding to single and double-capped proteasomes (19S-20S and 19S-20S-19S, respectively) and confirmed the higher amount of these complexes in patient cells (Figure 29C). As shown in Figure 3C, staining for PA28- $\alpha$ , a subunit of the alternative proteasome regulator PA28, revealed one major band corresponding to the position of the 20S proteasome, indicating the 20S proteasomes in these samples mainly consist of PA28-20S complexes. Again, patient fibroblasts exhibited a stronger expression level of such homo-PA28 complexes than their wild-type counterparts. To validate the notion that the homozygous pathogenic variant results in increased assembly of newly synthesized proteasome complexes, patient fibroblasts were compared to control ones for their content in various proteasome subunits using SDS-PAGE followed by western-blotting. As illustrated in Figure 3D, the steady-state expression level of most of the  $\beta$ ,  $\alpha$  and Rpt subunits was substantially higher in the patient's cells. Unexpectedly, the increased proteasome content was accompanied by a parallel rise of ubiquitin-modified proteins in these cells, suggesting that proteasomes from patients bearing the homozygous *PSMC3* pathogenic variant, although being in greater number, are ineffective. Altogether, these results point to a defective proteasome function in subjects carrying the deep intronic homozygous *PSMC3* pathogenic variant, which seems to be compensated by an ongoing assembly of newly synthesized 20S and 26S complexes.



**Figure 29. Fibroblasts derived from patient carrying the c.1127+337A>G homozygous *PSMC3* variation exhibit an increased amount of both proteasome complexes and ubiquitin-protein conjugates.**

**(A)** Whole-cell lysates from control and patient (case index, CI) fibroblasts were assessed by 3-12 % native-PAGE gradient gels with proteasome bands (30S, 26S and 20S complexes) visualized by their ability to cleave the Suc-LLVY-AMC fluorogenic peptide. **(B)** Ten micrograms of control and patient cell lysates were tested for their chymotrypsin-like activity by incubating them with 0.1 mM of the Suc-LLVY-AMC substrate at 37 °C over a 180-h period of time in quadruplicates on a 96-well plate. Indicated on the y-axis are the raw fluorescence values measured by a microplate reader and reflecting the AMC cleavage from the peptide. **(C)** Proteasome complexes from control and patient fibroblasts separated by native-PAGE were subjected to western-blotting using antibodies specific for Alpha6, Rpt5 (PSMC3) and PA28- $\alpha$ , as indicated. **(D)** Proteins extracted from control and CI PSMC3 were separated by 10 or 12.5 % SDS-PAGE prior to western-blotting using primary antibodies directed against ubiquitin and several proteasome subunits and/or components including  $\alpha$ 6,  $\beta$ 1,  $\beta$ 2,  $\beta$ 5,  $\beta$ 5i, Rpt2 (PSMC1), Rpt5 (PSMC3), Rpt3 (PSMC4), Rpt4 (PSMC6) and PA28- $\alpha$ , as indicated. For the PSMC3 staining, two exposure times are shown. Equal protein loading between samples was ensured by probing the membrane with an anti- $\alpha$ -Tubulin antibody.

We next sought to determine the impact of this variant on the ability of the cells to respond to perturbed protein homeostasis following proteasome dysfunction. To this end, both control and patient cells were subjected to a 16-h treatment with the  $\beta$ 5/ $\beta$ 5i-specific inhibitor carfilzomib in a non-toxic concentration prior to SDS-PAGE and western-blotting analysis using various antibodies specific for proteasome subunits. As shown in Figure 30, control cells exposed to carfilzomib could successfully compensate the applied proteotoxic stress by increasing their pools of intracellular proteasomes, as evidenced by elevated expression of all investigated  $\beta$ - and Rpt-subunits. As expected, this process was preceded by the processing of the ER membrane-resident protein TCF11/Nrf1 (Figure 30A), which is



**Figure 30.** Patient fibroblasts carrying the c.1127+337A>G homozygous *PSMC3* variation exhibit an exhausted TCF11/Nrf1 processing pathway which prevents them to upregulate proteasome subunits in response to proteotoxic stress.

**(A)** Control and patient (case index, CI PSMC3) fibroblasts were exposed to a 16-h treatment with 30 nM of the proteasome inhibitor carfilzomib or left untreated (as a negative control). Following treatment, cells were collected and subjected to RIPA-mediated protein extraction prior to SDS-PAGE and subsequent western-blotting using antibodies specific for ubiquitin, TCF11/Nrf1, Rpt1 (PSMC2), Rpt3 (PSMC4), Rpt5 (PSMC3), Rpt6 (PSMC5),  $\beta$ 1,  $\beta$ 2,  $\beta$ 5,  $\beta$ 5i,  $\beta$ 1i, PA28- $\alpha$  and  $\alpha$ -Tubulin (loading control) as indicated. For the TCF11/Nrf1 staining, two exposure times are shown. **(B)** Schematic diagram depicting TCF11/Nrf1 processing pathway in response to proteasome dysfunction and/or proteotoxic stress in patient carrying the deep intronic homozygous *PSMC3* variation (left) and in healthy subjects (right). Under normal conditions, TCF11/Nrf1 is a short-lived ER-membrane resident protein, which is rapidly subjected to proteasome-mediated degradation following retro-translocation to the cytosol by ER-associated degradation machinery (ERAD). In case of proteasome dysfunction (i.e. proteotoxic stress), the half-life of TCF11/Nrf1 is prolonged and become then a substrate for the NGLY1 and DDI2, thereby giving rise to a C-terminal cleaved fragment that enters into the nucleus. After nuclear translocation, cleaved TCF11/Nrf1 associates with Maf and promotes the expression of proteasomes genes, so that protein homeostasis can be restored.

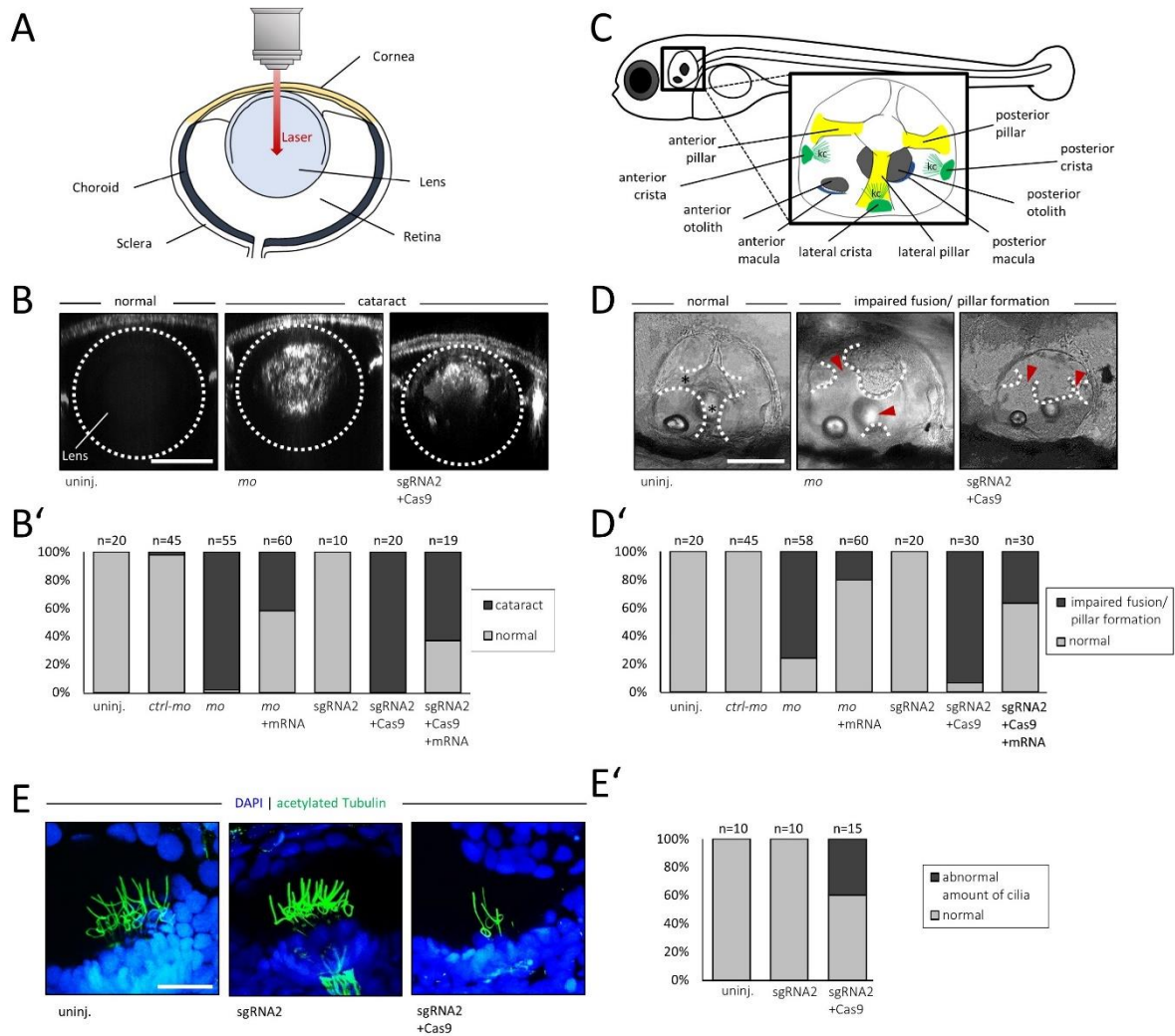
the transcription factor acting on nuclear genes encoding 19S and 20S proteasome subunits<sup>305–307</sup>. Strikingly, the level of processed TCF11/Nrf1 in response to carfilzomib was much lower in cells carrying the homozygous *PSMC3* pathogenic variant than that observed in control cells. Accordingly, the patient’s fibroblasts were unable to upregulate their proteasome subunits following proteasome inhibition, as determined by decreased expression levels of the  $\beta$ subunits (Figure 30A). Altogether these data indicate that, although exhibiting higher amounts of proteasomes under normal conditions, fibroblasts derived from patients carrying the homozygous *PSMC3* pathogenic variant fail to preserve protein homeostasis under stress conditions, as a result of proteasome deficiency.

**Effect of PSMC3 loss of function in zebrafish similar to patients' phenotype**

To establish a functional link between the observed decreased proteasome activity and the phenotype observed in the patients, we next assessed the lens and the ear in the zebrafish model. The zebrafish orthologue (Ensembl, ENSDARG0000007141) of human *PSMC3* is located on chromosome 7 with two predicted protein-coding splice variants (404 and 427 amino acids). Both zebrafish *psmc3* isoforms share 83 % sequence identity with the human orthologue. We confirmed by *in situ* hybridization that *psmc3* is maternally expressed and not spatially restricted (supplemental Figure S3.7).<sup>308</sup> Injection of morpholinos against *psmc3* generated zebrafish morphants embryos that were examined at 4 days post-fertilization (dpf) for lens or ear abnormalities. The lens size of morphants was slightly smaller than in control or uninjected embryos (supplemental Figure S3.8A). For cataract detection, we used a protocol based on confocal reflection microscopy, a labelling-free non-invasive imaging method that enables the detection of abnormal light reflection in the lens of living embryos (Figure 31A).<sup>149</sup> Morpholino injections revealed significant abnormal lens reflections in 95 % of the morphants (n=55), whereas only 2 % of 5 bp mismatched morpholino injected controls (n=45) and none of the uninjected embryos (n=20) showed cataract (Figure 31B-B').

As deafness has been reported for all three human patients, we subsequently examined the inner ear development of 4 dpf zebrafish morphants (Figure 31C). *psmc3* morphants displayed a smaller ear compared to control or uninjected embryos (supplemental Figure S3.8C). Interestingly, the majority of morphants presented anomalies during the semicircular canal morphogenesis. While the epithelium projections of all uninjected (n=20) and all control-injected (n=45) embryos were fused and formed pillars after 4 dpf, the canal projections of morphants failed to fuse in 79 % of the cases (n=58, Figure 5D-D'). The specificity of the morpholino was confirmed by a rescue experiment by co-injecting the full-length splice morpholino resistant zebrafish *psmc3* mRNA. The cataract phenotype was rescued in ~58 % of the cases (n=60) while the ear phenotype was rescued in 76 % of the cases (n=60) (Figure 31B'+D'). The outgrowth of the epithelial projections of the developing anterior and posterior semicircular canals begins around 48 hpf. From 57-68 hpf, the projections fuse in the center of the ear and form the pillars. This is followed by the outgrowth of the projection of the lateral semicircular canal around 57 hpf and completed by the fusion with the other two pillars in the center around 70 hpf.<sup>309</sup> To analyze the semicircular canal morphogenesis in *psmc3* morphants, life imaging was performed between 56 and 72 hpf in the transgenic line gSAIzGFF539A expressing a GFP signal in all three pillars. In all morphants (n=9), the projections failed to fuse and form pillars during the observed time frame. In two morphants (22 %), the outgrowth of epithelial projections even failed completely.





**Figure 31. *psmc3* morphants and F0 mosaic zebrafish exhibit cataract and show abnormalities during the semicircular canal development in the ear.**

**(A)** Scheme of a zebrafish eye. **(B-B')** Cataract detection revealed abnormal lens reflection in *psmc3* morpholino (MO)-mediated knockdown but not in controls (uninj, *ctrl-mo*). Similarly, abnormal lens reflection was also observed in embryos injected with sgRNA + Cas9 but not in sgRNA injected embryos without Cas9 (sgRNA2). Co-injection of wt *psmc3* mRNA with either *psmc3-mo* or sgRNA2 + Cas9 reduced the number of embryos presenting abnormal lens reflection. Scale bar = 50  $\mu$ m. **(B')** Quantification of embryos with abnormal lens reflection. **(C)** Representative image of a zebrafish ear at 4 dpf. kc= kinocilia. **(D-D')** bright field images of inner ear development (lateral position). **(D)** Epithelial projections were fused and formed canal pillars in 4-day-old uninjected and control injected fish (*ctrl-mo*, sgRNA2) but not in morphants (*mo*) and crispants (sgRNA2+Cas9). Co-injection of wt *psmc3* mRNA with *psmc3-mo* or sgRNA + Cas9 reduced the number of embryos presenting abnormal ear phenotype. Black asterisks indicate fused pillars. Red arrowheads mark unfused projections. Scale bar = 100  $\mu$ m. **(D')** Quantification of embryos with abnormal projection outgrowth. **(E)** An anti-acetylated tubulin antibody (green) staining revealed an abnormal amount of kinocilia in *psmc3* crispants (sgRNA2+Cas9) compared to uninjected and control injected embryos (sgRNA2). Nuclei are stained in blue with DAPI. Representative images show kinocilia of the lateral cristae. Scale bar = 20  $\mu$ m. **(E')** Quantification of embryos with an abnormal amount of kinocilia.

In contrast, the projections of uninjected (n=2) and control injected (n=2) embryos fused during the observed period in the center of the ear and formed canal pillars (Movie S3). Previous studies showed that a reduced number or a smaller size of otoliths, crystal-like structures required for the transmission of mechanical stimuli to the hair cells, can lead to deafness and balancing difficulties in zebrafish.<sup>310,311</sup> *psmc3* morphants did not present any otolith defect (supplemental Figure S3.9A-C). In addition, expression of *otopetrin*, a gene required for proper otolith formation, was unaffected at 28 hpf and 4 dpf (supplemental Figure S3.9D).

As autism has been reported for one of the patients and since brain malformation have been reported previously in some autistic patients and autistic zebrafish morphants<sup>312</sup>, we investigated possible morphological brain changes in *psmc3* morphants. *In-situ* hybridization targeting brain markers such as *krox20*, *msxc*, *her8a* and *sox19b* was performed on 24 hpf embryos but did not reveal any obvious differences in their expression patterns (supplemental Figure S3.10).

To confirm these results, we additionally used the CRISPR/Cas9 system to knockdown *psmc3* (Figure 31B-D). The high cutting efficiency of the CRISPR/Cas9 founders (F0) were evaluated to 55.2 % (gRNA1) and 53.7% (gRNA2) (supplemental Figure S13).<sup>224</sup> CRISPR/Cas9 founders (i.e. crispants) are often genetically mosaics. However, in cases of highly efficient gRNAs, they have been shown to recapitulate mutant phenotypes successfully.<sup>221–223</sup> Both *psmc3* crispants used in this study displayed a cataract phenotype (100% of gRNA1+Cas9 (n=20) and 95 % of gRNA2+Cas9 (n=20), whereas none of the control embryos (injected with gRNA1 (n=20) or gRNA2 (n=10) without the Cas9 protein) showed abnormal lens reflections (Figure 31B-B' and supplemental Figure S3.11A-A'). Moreover, we recapitulated the ear phenotype seen in *psmc3* morphants. While the ear pillars of uninjected (n=10) and control-injected (n=20) embryos formed after 4 dpf, the projections of both crispants failed to fuse in 65 % of gRNA1 (n=10) and 93% of gRNA2 (n= 30) injected embryos (Figure 31D-D' and supplemental Figure S3.11B-B'). Co-injection of gRNA2, Cas9 and the *psmc3* mRNA also led to a partial rescue of the lens and ear phenotype with only 63 % of abnormal lens reflection instead of 100 % (n=19) and 37 % of unfused canal projections instead of 93 % (n=30) (Figure 31D'). Performing an *in situ* hybridization examining the mRNA expression of *versican a* and *versican b*, two genes suggested to be required for a proper canal fusion event<sup>309</sup>, significant differences could be observed after 72 hpf. Indeed, both genes were highly expressed in the whole ear tissue, whereas *versican a* was not expressed and *versican b* was restricted to the dorsolateral septum in wildtype or control injected embryos (supplemental Figure S3.12B-D).

As kinocilia, specialized microtubule-based structures found on the surface of hair cells, have been shown to play a key role at least in mechanosensation during development<sup>313</sup>, we immunostained crispants (sgRNA2 injected with Cas9) and control injected embryos (sgRNA2 without Cas9) at 5 dpf

using anti-acetylated tubulin antibody. In 40 % of crispants (n=15), a reduced number of kinocilia was observed while the number of control injected embryos (n=10) was similar to that of uninjected embryos (n=10) (Figure 31E+E').

Taken together, these zebrafish assays confirm that *psmc3* plays a very important role in the development of a transparent lens and the semicircular canals of the inner ear - reminiscent of the human phenotype described herein.

#### 4.3.4 Discussion - Project 3

In this study, we describe a novel homozygous deep intronic splice variant, identified in 3 patients with an unusual neurosensorial disease combining early onset deafness, cataracts and sub cutaneous deposits, in *PSMC3* encoding one of the proteasome subunits. Clinical data and functional analysis in patient's cells and zebrafish proved the effect of the variation and the consequences leading to a recessive form of proteasome deficit with a haploinsufficiency mechanism. The observation of a neurosensory disease broadens the spectrum of ubiquitin-proteasome system (UPS) related disorders. Sequencing the entire genome of patients gives access to the whole spectrum of their variations and possibly disease-causing ones. WGS is a powerful tool<sup>314</sup> helping to identify variations not covered or missed by WES such as structural variations<sup>186</sup> or deep intronic variations<sup>315</sup>. Interestingly, in this study, we combined to WGS, homozygosity mapping and *in silico* predicted interactors to narrow down to the region of *PSMC3*. The 3 patients carried an homozygous deep intronic variation (i.e. >100 bases from the exon-introns boundaries)<sup>315</sup> with a predicted splicing effect on the *PSMC3* gene that we confirmed on the patient's cells. We then focused on demonstrating the effect of this variation at the level of the proteasome.

The 26S proteasome, the chief site of protein turnover in eukaryotic cells, consists of 2 complexes: a catalytic 28-subunit barrel shaped core particle (20S) that is capped at the top or the bottom by one 19 subunit regulatory particle (19S). The core particle contains the catalytic subunits  $\beta 1$ ,  $\beta 2$  and  $\beta 5$  exhibiting caspase-, trypsin- and chymotrypsin-like activities, respectively. Recognition of a substrate with the requisite number and configuration of ubiquitin is mediated principally by both Rpn10 and Rpn13 subunits which act as ubiquitin receptors<sup>316,317</sup>. To allow substrate degradation, ubiquitin is first removed by Rpn11, a metalloprotease subunit in the lid<sup>318</sup>. The globular domains of a substrate are then unfolded mechanically by a ring-like heterohexameric complex consisting of six distinct subunits, Rpt1 to Rpt6, which belong to the ATPases-associated-with-diverse-cellular-activities (AAA) family.<sup>319</sup> *PSMC3* encodes for Rpt5 involved in the substrate unfolding and translocation, which are then presumably catalyzed.<sup>304,320</sup>

To our knowledge, this is the first report of a human biallelic pathogenic variant occurring in one of the ATPase Rpt subunits of the base of the 19S regulatory particle. Recently, *de novo* pathogenic variations in the non-ATPase subunit *PSMD12* (Rpn5) of the 19S regulator lid of the 26S complex have been reported in 6 patients with neurodevelopmental disorders including mainly ID, congenital malformations, ophthalmic anomalies (no cataracts), feeding difficulties, deafness (unspecified type for 2 patients/6), and subtle facial features.<sup>222</sup> *PSMD12* variants have been also associated to a large family with ID and autism and 1 simplex case with periventricular nodular heterotopia.<sup>321</sup> *PSMD12* is highly intolerant to LoF variations and the most likely effect is haploinsufficiency due to the *de novo*

heterozygote occurrence of loss of function truncating, non-sense or deletion variants. Interestingly, according to the gnomAD and DDD data<sup>322,323</sup>, *PSMC3* is also predicted to be extremely intolerant to LoF variation. Indeed, the haploinsufficiency score of *PSMC3* is 4.76 that is within the high ranked genes (e.g. HI ranges from 0 % to 10 %) from the DECIPHER data. The pLI score (0.96) makes it among the highest intolerant genes (e.g. a score >0.9 defines the highest range) with only 3 observed LoF variants versus 23.2 predicted and confidence interval = 0.13). This explains also maybe the rarity of LoF variations found to date in this gene. In our cases, this is the first time that biallelic class 5 variations are reported in one of the proteasome subunits delineating a recessive mode of inheritance. Nevertheless, the fact that the homozygous variation is affecting the splicing machinery and leads to a reduced but not abolished expression of *PSMC3* could mimic a possible haploinsufficiency mechanism.

Both proteomic and biochemical approaches undertaken in this study revealed that the deep intronic homozygous *PSMC3* variation is associated with increased amounts of 26S and 20S-PA28 proteasome complexes (Figures 28C and 29). The observation that patient cells concomitantly increase their intracellular pool of ubiquitin-protein conjugates (Figures 28A and 29D) is surprising and strongly suggests that such proteasomes are defective. In support of this notion, we found that, although carrying greater amounts of proteasomes, patient fibroblasts did not exhibit higher chymotrypsin-like activity compared to control cells, which can be at least partly explained by upregulation of PA28 and the concomitant increased peptide-hydrolysis<sup>324</sup> (Figures 29A and B). These data led us to conclude that the increased steady-state expression level of the proteasome subunits observed in patient fibroblasts might reflect a constitutive *de novo* synthesis of proteasomes, which aims to compensate the dysfunctional ones. In mammalian cells, a major compensation mechanism for proteasome dysfunction is governed by the ER membrane-resident TCF11/Nrf1 protein<sup>305–307</sup>. Typical stimuli for TCF11/Nrf1 activation include proteasome inhibition and/or impairment which results in the release of C-terminal processed TCF11/Nrf1 fragment from the ER membrane following a complex series of molecular events involving the enzymes NGLY1 and DDI2 (Figure 30B). The cleaved TCF11/Nrf1 fragment enters then into the nucleus and acts as a transcription factor to promote the expression of ARE-responsive genes including 19S and 20S proteasome subunits, thereby augmenting the pool of proteasomes so that protein homeostasis can be preserved.<sup>305–307</sup> Strikingly and in contrast to control cells, TCF11/Nrf1 is constitutively processed in patient cells (Figure 30A), confirming that patients' proteasomes were impaired. This, in turn, gives rise to a pathological vicious circle of events in which TCF11/Nrf1 and defective proteasomes stimulate each other. We reasoned that such a process may result in a pathway overload which in turn reduce the ability of TCF11/Nrf1 to respond to further proteotoxic stress. Consistent with this hypothesis and unlike control cells, patient fibroblasts were not capable of upregulating their proteasome subunits when challenged with proteasome inhibitor carfilzomib (Figure 4A). This result is of great importance, as it demonstrates that, under stress

conditions, cells carrying the deep intronic homozygous *PSMC3* variation suffer from proteasome deficiency. Because cataract and semicircular canal malformations are observed in zebrafish embryos depleted with *PSMC3* -and *a fortiori* proteasomes-, these data established a clear cause-and-effect relationship between the deep intronic *PSMC3* variant and the acquisition of patients' phenotype. On the other hand, one cannot exclude that the pathogenesis of the homozygous *PSMC3* variation may involve additional mechanisms. Because target genes of TCF11/Nrf1 include anti-inflammatory factors<sup>325,326</sup>, it is also conceivable that inflammation might play a role in this process. This assumption would be in line with the observation that subjects suffering from other loss-of-function proteasome variations such as *PSMB8*<sup>327–329</sup>, *PSMA3*, *PSMB4*, *PSMB9*<sup>330</sup> and/or *POMP*<sup>331</sup> exhibit an inflammatory phenotype including joint contractures. In any case, the potential contribution of innate immunity to the pathogenesis of the homozygous *PSMC3* variant via TCF11/Nrf1 warrants further investigation<sup>327,329</sup>.

The ubiquitin-proteasome system (UPS) is a protein degradation pathway that regulates the intracellular level of proteins involved in a very wide variety of eukaryote cellular functions. Thus, it is not surprising that this pathway is related to multiple human conditions including cataract, where an overburden of malfunctioning and aggregated proteins cannot be adequately removed by the UPS<sup>332</sup>. Moreover, protein degradation dysfunction is recognized as a widespread cause of neurodegenerative diseases such as Parkinson, Huntington or Alzheimer diseases. Several inherited rare disorders have been shown to be related to directly UPS dysfunction in enzymes of the ubiquitin conjugation machinery: *UBE3A* in Angelman syndrome<sup>333</sup>, *UBE2A* in X-linked syndromic ID<sup>334</sup>, *UBE3B* in Kaufmann oculo cerebrofacial syndrome (classified as blepharophimosis – mental retardation syndrome) or other related enzymes such as *HUWE1* an ubiquitin ligase in X-linked for dominant ID syndrome<sup>335 336</sup> and *OTUD6B* a deubiquitinating enzyme for neurodevelopmental disability with seizures and dysmorphic features<sup>337</sup>. The phenotypic tropism for neurodevelopmental disorders especially with intellectual disability and other brain dysfunctions (autism, seizures, brain malformations) seem to be overall a hallmark of the gene alterations related with the UPS inherited dysfunctions. In the family presented herein, all 3 patients had various neurodevelopmental anomalies (autism, ataxia, mild ID) and mild facial dysmorphism.

However, the striking clinical presentation of the cases reported herein is the combination of early onset deafness and early onset cataracts that motivated (independently) the referral of the 3 cases. These features have not yet been related to this group of disorders to date. Moreover this association has only been reported twice in larger syndromic forms such as the dominant form of *WFS1* but in the context of neonatal/infancy -onset diabetes<sup>301</sup> and the Aymé-Gripp syndrome (including also intellectual disability, seizures and Down syndrome like facies) with *de novo* pathogenic variants in the

*MAF* gene<sup>300</sup>. Variations in the later one were shown to impair *in vitro* MAF phosphorylation, ubiquitination and proteasomal degradation.

The zebrafish model was extremely useful to demonstrate the effect of a reduced *psmc3* expression leading to inner ear anomalies and lens opacities. The overall development of the lens and inner ear as well as the underlying gene regulatory networks are highly conserved throughout evolution.<sup>70</sup> Pathogenic variations or knockdown of genes implicated in human cataract and/or deafness in zebrafish have been shown to recapitulate similar phenotypes.<sup>146–150</sup> Using two strategies (morpholino and CRISPR/Cas9) the fish developed a cataract and ear phenotype that could be rescued. We did not observe other obvious anomalies. Interestingly, a reduction of proteasome activity has previously been associated with lens defects in zebrafish. The knock out of the zebrafish gene *psmd6* and the knock down of *psmd6* and *psmc2*, both encoding for proteins of the UPS, resulted in a severe impairment of lens fibre development. The ear phenotype of the *psmd6* mutant and both morphants was not assessed.<sup>338</sup> A direct link between the UPS and auditory hair cell death or impaired semicircular canal morphogenesis has not been described in zebrafish yet. However, knock down of *atoh1*, a gene regulated by the UPS, has been shown to severely affect hair cell development in the inner ear of zebrafish.<sup>339</sup> The of canal pillars observed in zebrafish *psmc3* morphants and crispants might be also a secondary effect, as abnormal sensory cristae with few hair cells have been previously assumed to lead to an abnormal development of semicircular canals.<sup>340,341</sup>

In conclusion, our work demonstrates the implication of a deep intronic variant in a novel ultra-rare neurosensorial syndrome with early onset cataract and deafness in one of the proteasome subunits, *PSMC3*. Although, *de novo* dominant variations have been associated to several proteasome related disorders, we report for the first time a bi allelic pathogenic variant. Our observations strongly suggest that the amount of PSMC3 is critically implied in the development and maintenance of the inner ear and the lens.

### 4.3.5 Supplementary Information – Project 3 (S3)

#### **Authors contributions**

E.S., S.S., F.S., V.P., C.S.S., V.L., D.L. and H.D. gathered data from patients and performed clinical investigations; S.M., A.B., J.F.D., I.P., V.G. and K.C. gathered sequencing data and performed analyses; **A.KH.**, F.E., C.S., S.B., B.A.Z. and S.F. performed cell biology experiments and data analyses; C.S., L.K., J.C. and P.H. designed and performed the mass spectrometry experiments and data analyses; **A.KH.**, M.T., K.K., C.E. and U.S. designed and performed the zebrafish experiments and data analyses; **A. KH.**, J.M., U.W. and H.D. analyzed the data and wrote the paper; F.E., V.G., E.S., S.S., S.F., S.B., C.S., U.S., P.H., C.E. and EK contributed to manuscript writing; J.M., U.W. and H.D. provided direction for the project, conceived and designed the experiments.

#### **Supplementary methods**

##### **Method S3.1 Whole Exome Sequencing analysis**

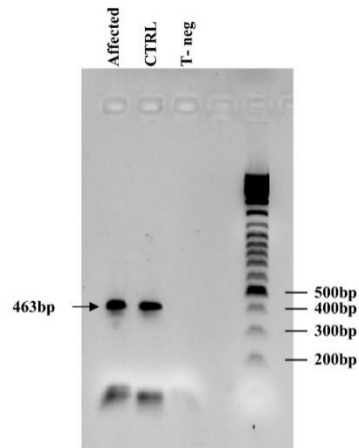
From 96,222 to 102,072 genetic variants (SNV/indel/SV) were identified per proband from the WES analysis (supplemental Table S3.1). Bioinformatic analyses highlighted a unique homozygous missense variation in the *DGKZ* gene (NM\_001105540.1:c.1834G>A, p.Ala612Thr) encoding for the Diacylglycerol Kinase Zeta, located in a homozygous region of interest on chromosome 11 (supplemental Figure S3.1). This variant, reported previously with a gnomAD frequency of 0.018 %, was predicted tolerated by SIFT<sup>278</sup> and neutral by PolyPhen-2<sup>279</sup> and was finally manually ruled out, as we were unable to explain the patients phenotype based on the gene function.

##### **Method S3.2 Whole Genome Sequencing analysis**

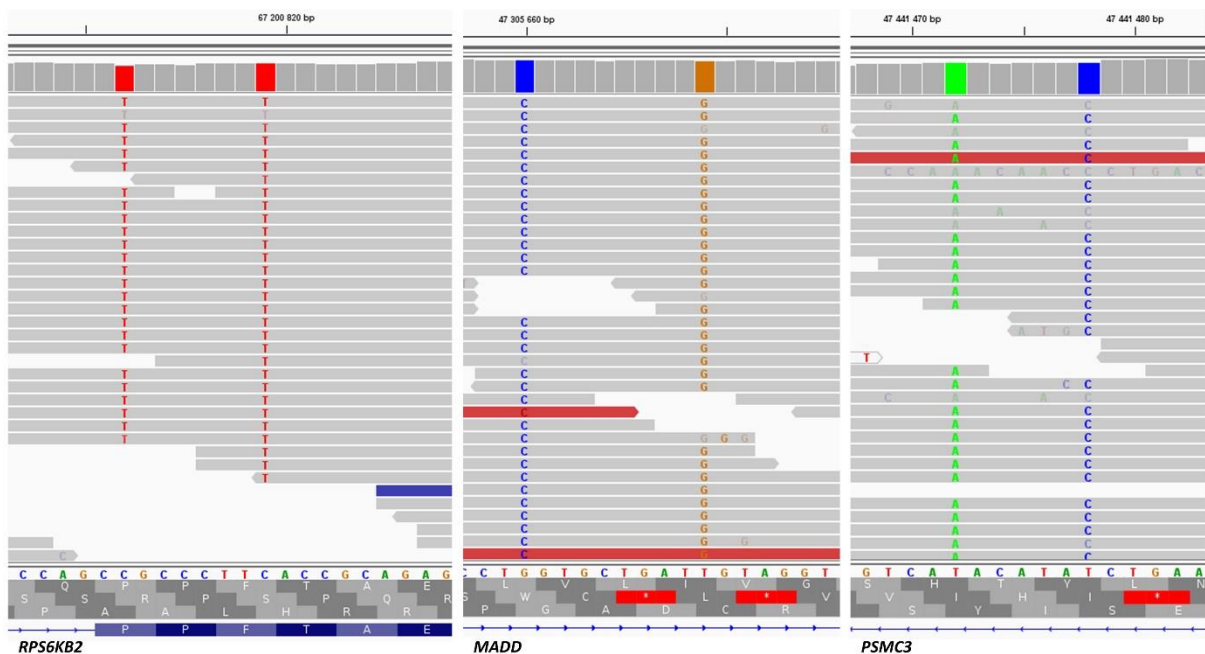
From 4,990,312 to 5,165,496 genetic variants (SNV/indel/SV) were identified per individual from the WGS analysis (Table S2). Bioinformatic analyses (see Materials and methods) highlighted 6 homozygous variations in 5 genes (*ATG13*, *CELF1*, *MADD*, *PSMC3* and *RPS6KB2*).

Among those, the variation in *ATG13* (chr11(GRCh37):g.46680484A>G, NM\_001346315.1:c.696-458A>G) is a rare (0.032 % maximum allele frequency) deep intronic variant predicted to create a potential acceptor splice site. RNA analysis of *ATG13* (RT-PCR on skin fibroblast from individual II.4 see Materials and Methods) surrounding intron 12 did not reveal any aberrant splicing event leaving this variant out (see below the results of the RT-PCR amplification between exon 9 and exon14/15, Table S5).



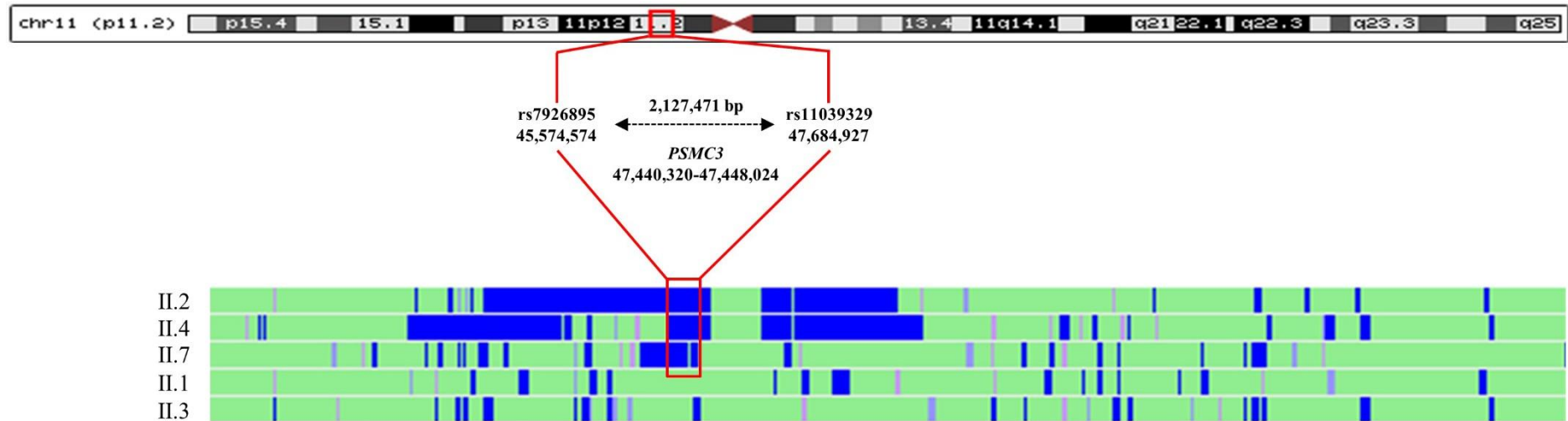


The *RPS6KB2* and *MADD* variations corresponded to 2 likely false positive calls. The *RPS6KB2* variant (chr11(GRCh37):g.67200812\_67200819delinsTGCCCTTT) was localized at 2 highly frequent SNP: rs55987642 and rs4930427 (maximum allele frequency in gnomAD >6 %). The same applies to the *MADD* variant (chr11(GRCh37):g.47305660\_47305669delinsCGTGCTGATG) with rs12573962 and rs3816725 (maximum allele frequency in gnomAD >9 %). This apply also to another variation in *PSMC3* (chr11(GRCh37):g.47441472\_47441478delinsAACATAC). This can be visualized on the IGV<sup>191</sup> view below.



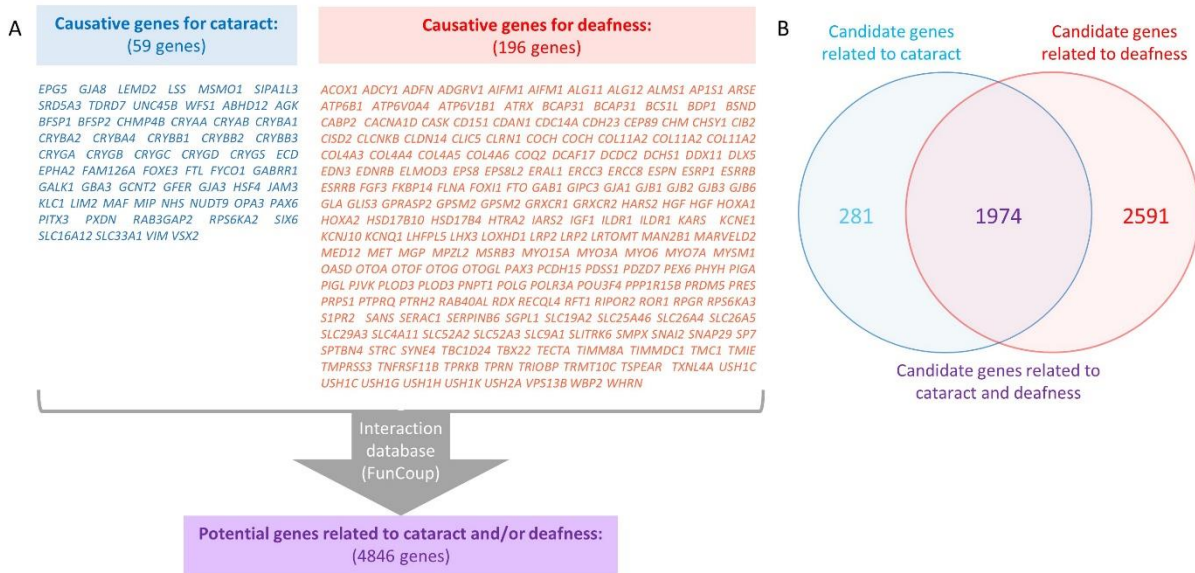
The variation in *CELF1* (Chr11(GRCh37):g.47489405T>C, NM\_001330272.1:c.\*4289A>G) is localized in the 3' UTR and no prediction could be associated to it, leaving this candidate out.

## Supplementary Figures



**Figure S3.1. Analysis of homozygosity region in chromosome 11.**

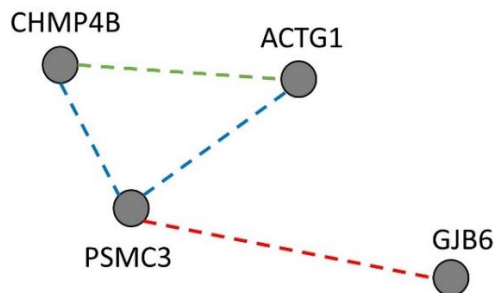
Homozygosity regions for three affected individuals (II.2, II.4, and II.7) and two unaffected siblings (II.1 and II.3) were analyzed using the HomoSNP software. The areas of homozygosity with >25 SNPs are colored in blue, whereas homozygosity regions defined by 15–25 consecutive SNPs are colored in pink. The 3 affected individuals share a common homozygous region between rs7926895 and rs11039329 resulting in a 2.12 Mb region encompassing the *PSMC3* gene.



**Figure S3.2. Selection of candidate genes related to deafness and/or cataract.**

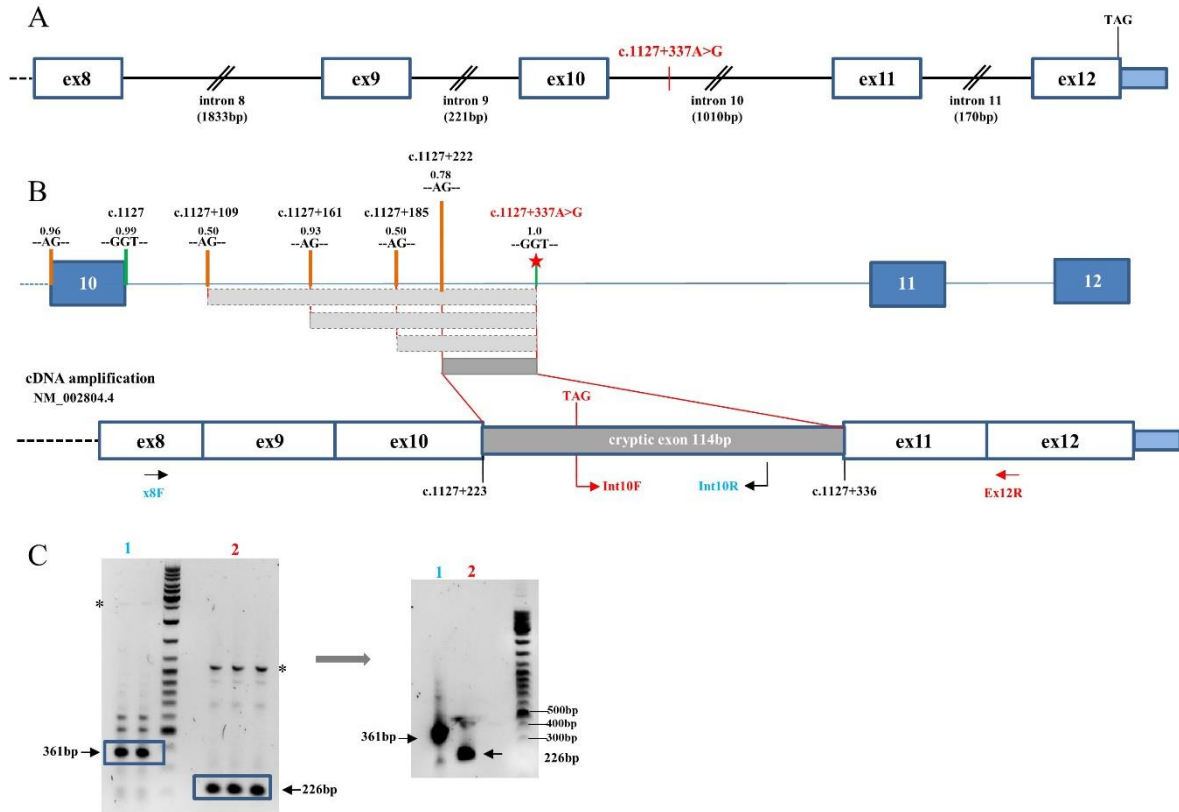
**(A)** Based on the OMIM database<sup>342</sup> and careful literature review two high-confidence reference gene sets including either isolated or syndromic cataract or deafness genes were defined. This resulted in 59 known cataract genes and 196 known deafness genes. Nevertheless, as the patients have congenital disorder, we excluded genes associated with ocular anomalies in which the cataract is a complication of the disease. These reference lists were used to query for potential interacting genes in the FunCoup database (4.0)<sup>343</sup>. Applying a confidence threshold of 0.8 and a single level of interaction a list of 4846 candidate genes related to cataract and/or deafness genes was defined.

**(B)** The 4846 candidate genes can be divided into 2255 candidate genes related to cataract and 4565 candidate genes related to deafness of which 1974 were in common.



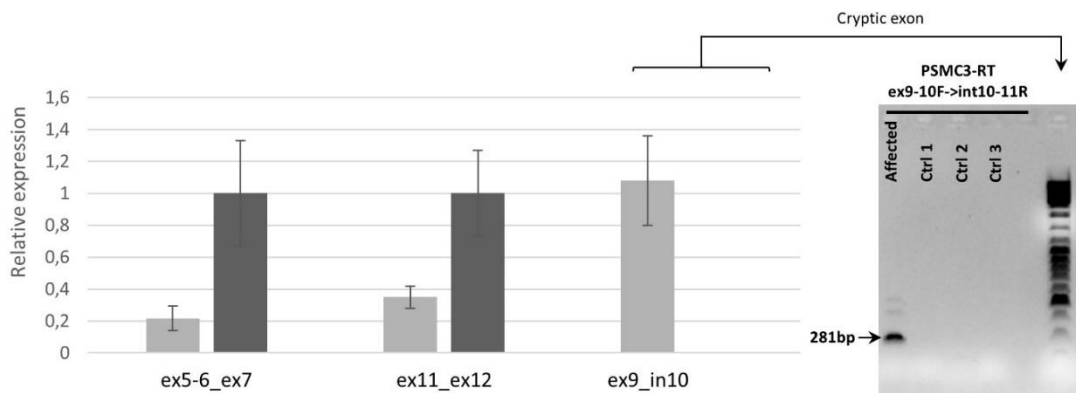
**Figure S3.3. Types of interactions between CHMP4B, ACTG1, GJB6 and PSMC3.**

Functional couplings displayed by FunCoup<sup>11</sup> are represented here with a blue line for protein-protein interaction experiment (Yeast2Hybrid), a red line for complex co-membership and a green line for a co-membership in a metabolic pathway. They are detailed in Table S4.



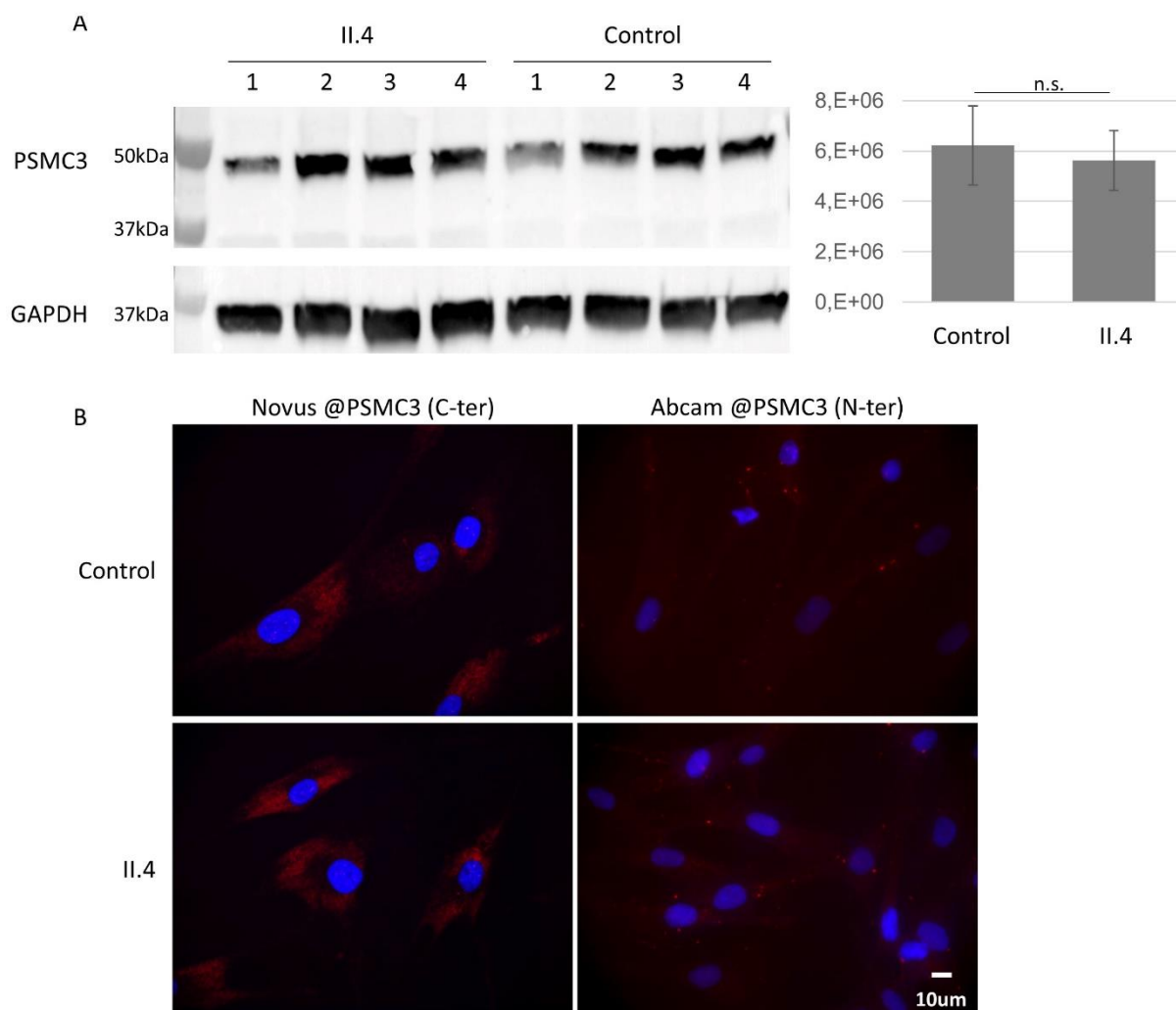
**Figure S3.4. Identification of the boundaries of the inserted cryptic exon.**

(A) Position of the deep intronic mutation on the *PSMC3* gene creating a strong cryptic splice donor site. Positions are given according to NM\_002804.4 and positioned on the chromosome 11: 47,438,320-47,445,024 (11,075 bp). (B) Schematic representation of the potential acceptor sites determined by the highest NNSPLICE score (shown below the HGVS nomenclature) as well as the resulting cryptic exons are indicated. The acceptor site at position c.1127+222 resulting in the cryptic exon displayed in dark grey was proven experimentally. (C) Identification of the acceptor site by RT-PCR using two overlapping pairs of primers, ex8F-int10R (1) and int10F-ex12R (2). Bands of the expected sizes (361bp and 226bp, blue squares) were seen and cut for PCR reamplification (right gel) and subsequent sequencing. \*indicates amplification of genomic DNA (bands at 2669bp and 1130bp).



**Figure S3.5. *PSMC3* mRNA quantification.**

mRNA quantification between exons 5-7 and 11-12 show a significant 40/50 % decrease on both fragments in the patient (II.4) as shown in light grey compared to the controls as shown in dark grey. The reference used is GAPDH. mRNA of the cryptic exon amplified using primers in exon 9 and in intron 10 was present at a high level in the patient while it was absent in the controls. The agarose electrophoresis shows the band corresponding to the cryptic exon at 281bp as amplified from the cDNA of the patient. Three independent controls of unaffected individuals were used.

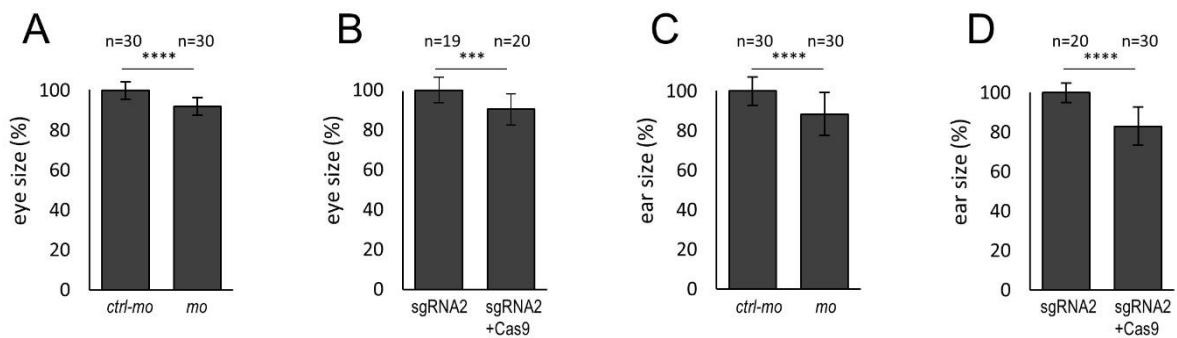


**Figure S3.6. No difference in PSMC3 expression or localization.**

**(A)** PSMC3 expression is comparable in controls and affected fibroblasts. Fibroblasts from four control individuals and from the patient II.4 (four separate batches of cells) were collected and PSMC3 detected by western blot in the whole cell lysate (Abcam antibody directed against the N-terminal thus recognizing both the full length and truncated forms of the protein). The secondary antibody was goat anti-mouse Alexa Fluor coupled 568 IgG (Invitrogen). GAPDH was used as a loading control and quantification of PSMC3 relative to GAPDH performed using Image Lab. n.s.: non-significant. **(B)** No difference in PSMC3 localization: Immunofluorescence of control and affected fibroblasts (II.4) labeled with antibodies directed against the N- or C-terminal part of PSMC3 and stained with DAPI.

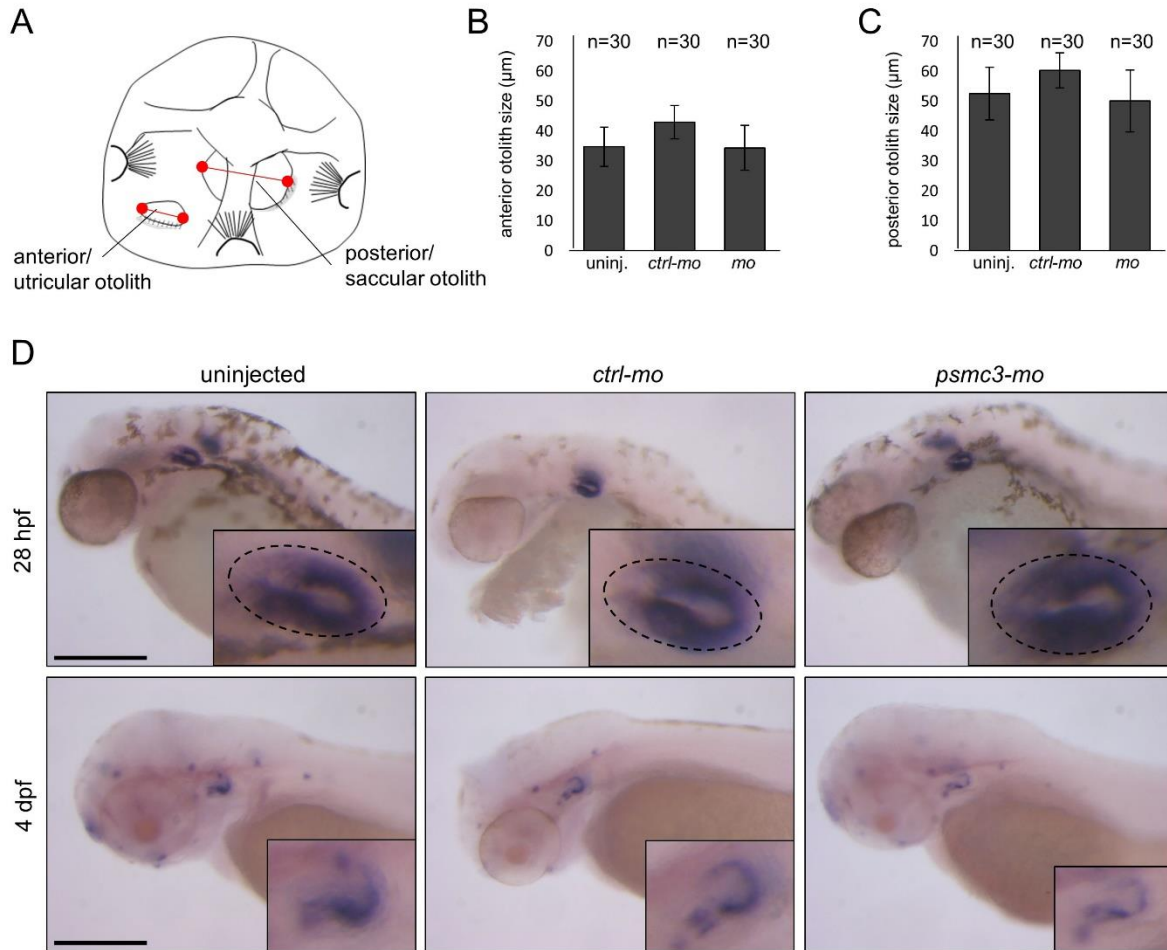


**Figure S3.7.** *In situ* hybridization with *psmc3* antisense probe showing ubiquitous expression (4 hpf-48 hpf). Red arrows indicate the zebrafish ear. Scale bars = 250  $\mu$ m.



**Figure S3.8.** *psmc3* morphants and crispants exhibit smaller lenses and inner ears.

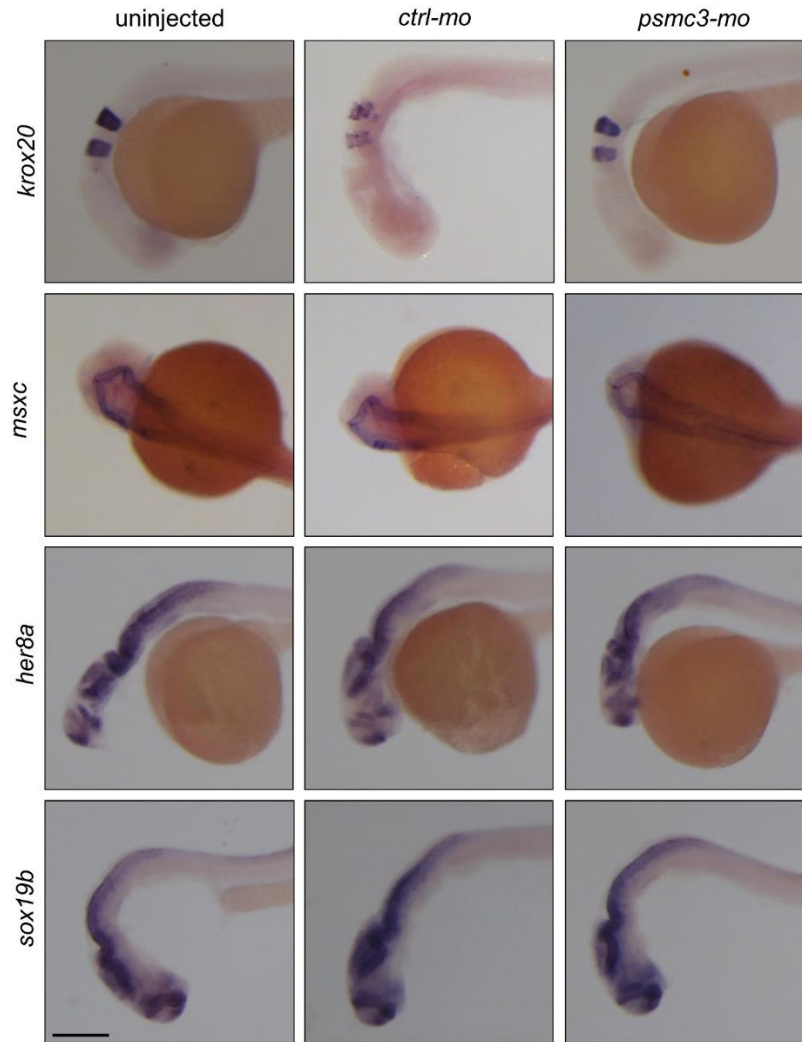
**(A-B)** Size quantification of the lens in morphants (*mo*) and F0 mutants targeted with sgRNA2 at 4 dpf. Morphants (*mo*) and crispants (sgRNA2 + Cas9) have slightly smaller lenses compared to uninjected and control injected embryos (*ctrl-mo*, sgRNA2). **(C-D)** Size quantification of the inner ear in morphants (*mo*) and crispants (sgRNA2) at 4 dpf. Mo- and sgRNA2 + Cas9 injected embryos present a smaller inner ear phenotype compared to uninjected and control injected embryos (*ctrl-mo*, sgRNA2). Statistical significance was determined using the unpaired, t Test, ns = non-significant. \*  $p < 0.05$ . \*\*  $p < 0.01$ . \*\*\*  $p < 0.001$ . \*\*\*\*  $p < 0.0001$ . Significance is determined relative to control injected embryos.



**Figure S3.9. Otolith development is not affected in *psmc3* morphants.**

(A) Representative image of a zebrafish ear at 4 dpf. For the evaluation of otolith size diameter both otoliths were measured (longest distance, shown in red). (B-C) Otolith sizes were measured and analyzed with ImageJ. (D) Lateral views of 28 hpf and 4 dpf old embryos after morpholino injection showing expression of *otopetrin*, a gene required for otolith formation. Scale bars = 250 µm.

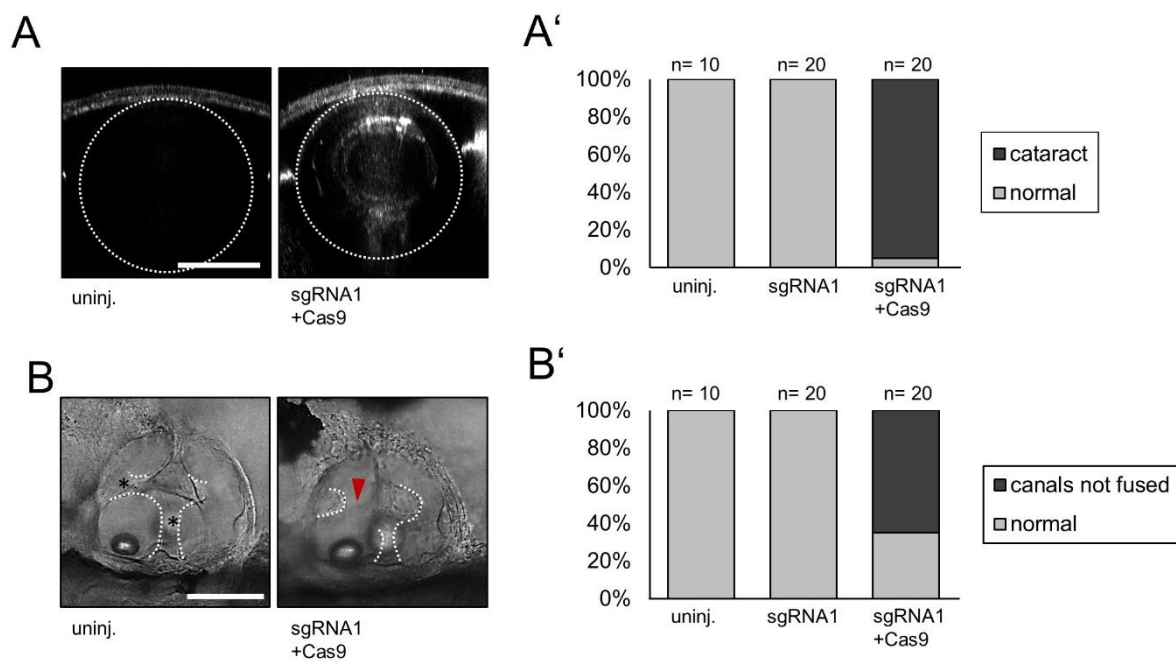




**Figure S3.10. *psmc3* morphants display no obvious brain malformations at 24 hpf.**

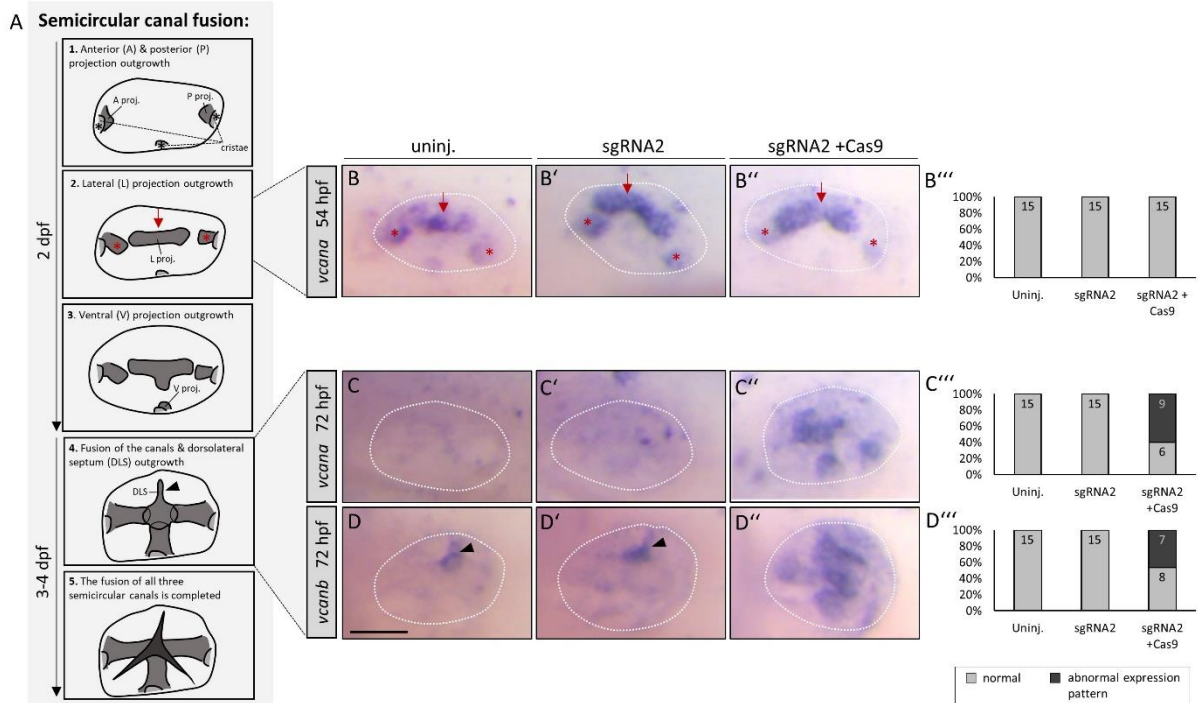
Lateral views showing expression of brain markers *krox20*, *msxc*, *her8a*, and *sox19b*. Scale bar = 250  $\mu$ m.





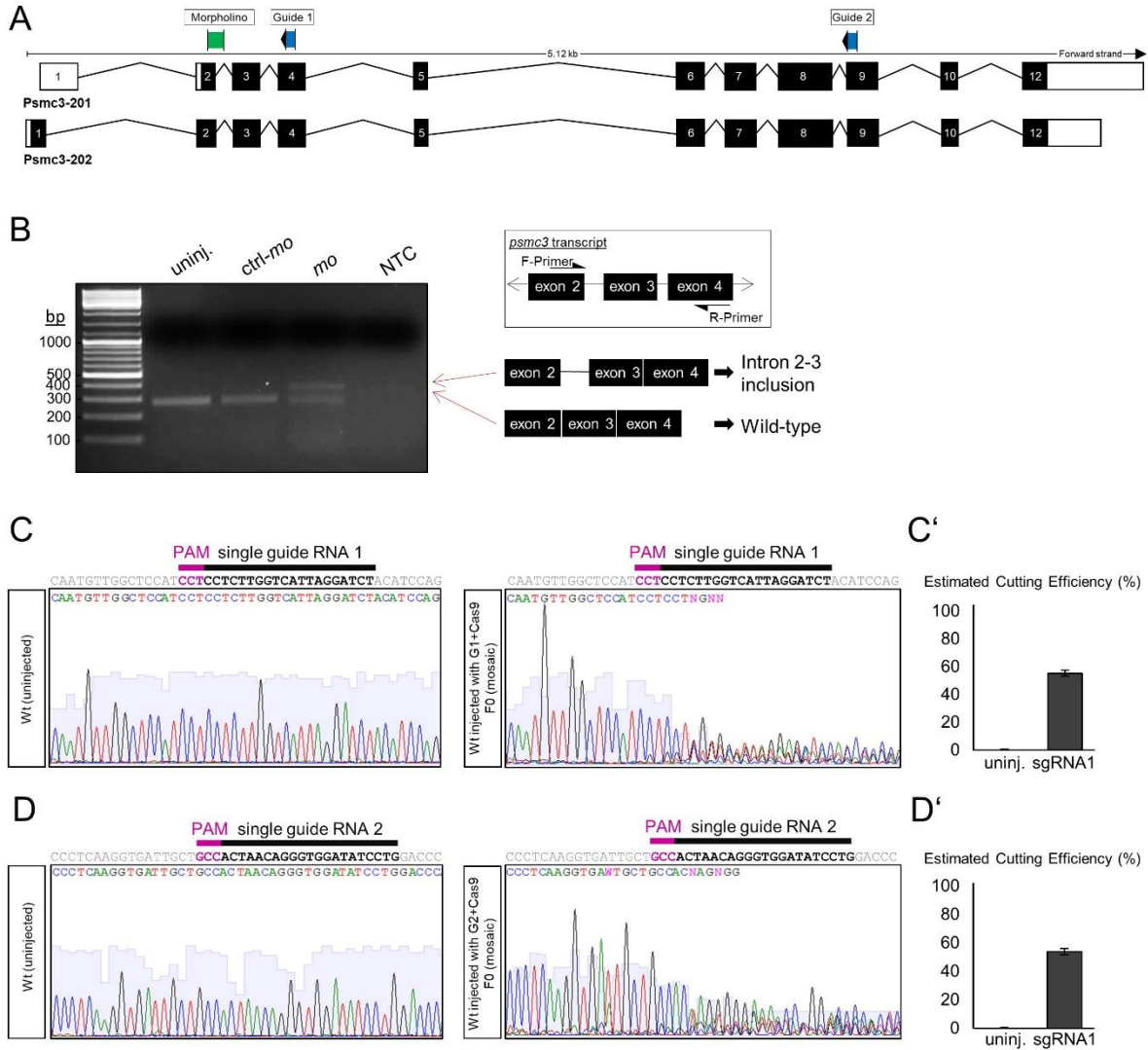
**Figure S3.11. A second guide RNA (sgRNA1) confirms the cataract and ear phenotype seen in morphants (*mo*) and crispants (sgRNA2).**

**(A)** Injection with sgRNA1 + Cas9 resulted in abnormal lens reflection but not in sgRNA injected embryos without Cas9 (sgRNA1) or uninjected embryos. Scale bar = 50  $\mu$ m. **(A')** Quantification of embryos with abnormal lens reflection. **(B)** Semicircular canals were fused in 4-day-old uninjected and sgRNA1 only injected fish but not in crispants (sgRNA1+Cas9). Black asterisks indicate fused pillars. Red arrowheads mark unfused projections. Scale bar = 100  $\mu$ m. **(B')** Quantification of embryos with an abnormal ear phenotype.



**Figure S3.12. Expression of genes involved in the zebrafish inner ear development in *psmc3* crispants.**

**(A)** Schematic representation of the semicircular canal morphogenesis between 2 and 4 dpf (inspired by Geng et al., 2013).<sup>309</sup> **(B-B''')** The expression of *versican a* (*vcana*) in *psmc3* crispants is normal after 54 hpf, but is **(C-C''')** clearly upregulated after 72 hpf compared to uninjected and control injected embryos. **(D-D''')** *Versican b* (*vcanb*) is expressed in the dorsolateral septum (black arrowhead), whereas in crispants it is highly upregulated in the whole unfused canal tissue. **(B''', C''', D''')** Quantification of embryos with an abnormal mRNA expression pattern. Scale bar = 50  $\mu$ m.



**Figure S3.13. Design and efficiency of morpholinos targeting *psmc3* pre-mRNA and guide RNAs targeting *psmc3*.**

**(A)** Simplified scheme of the 2 zebrafish *psmc3* isoforms. Morpholino target sequence is shown in green. Guide RNA target sequences are indicated in blue. **(B)** *psmc3-mo* injection affects splicing and leads to the inclusion of an intron. Regions amplified during PCR are shown with arrows. NTC: no template control. **(C-D)** Guide RNAs are able to cut multiple times independently during development leading to a set of small deletions and insertions. **(C)** Injection of 300 ng/μl guide RNA1 and Cas9 introduces indel in *psmc3* exon 4 with a cutting efficiency of 55.2%. **(D)** Injection of 300 ng/μl guide RNA2 and Cas9 introduces indel in *psmc3* exon 8 with a cutting efficiency of 53.7%. **(C'+D')** The web interface PCR-F-Seq q (<http://iai-gc-server.iai.kit.edu>) was used to quantify the cutting efficiency of both guide RNAs.<sup>224</sup>

**Table S3.1. Summary of the whole exome sequencing results.**

SNV: Single nucleotide variation, indel: gain or loss of up to 50 nucleotides at a single locus, SV: Structural Variation. Annotations are gathered using Alamut Batch v1.11, especially for the variation databases including gnomAD (v2.0.2, Oct. 2017), 1000 Genomes Project phase3 release (version 20150813 v5b) and the following predictions including phastCons (UCSC, 44 vertebrates). Effect on the splice has been evaluated using the MaxEntScan <sup>344</sup>, NNSPLICE 0.9 <sup>345</sup> and Splice Site Finder <sup>346</sup> by calculating score change between the wild type and the mutated sequences expressed as a percent differences. Missenses have been evaluated using default parameters from PolyPhen-2 (2.2.2) <sup>279</sup> and SIFT 4.0.3 <sup>347</sup>. Default cut-offs used have been described in VaRank <sup>348</sup> for both type of predictions. Exclusion of SV with a DGV (Gold standard from 20160515) frequency > 1 % is done only with studies of more than 1000 individuals.

	II.4			II.2			II.7			II.6		
	SNV	indel	SV	SNV	indel	SV	SNV	indel	SV	SNV	indel	SV
Total number of variants	88,136	13,923	13	86,637	13,009	12	84,019	12,698	22	83,948	12,253	21
After exclusion of variants with an allele frequency > 1% (gnomAD, 1000G, internal exome database, DGV)	9,198	1,274	6	7,993	1,138	5	7,638	1,090	11	7,667	1,099	5
After exclusion of SNV/indel found in the homozygous state in gnomAD and in our internal exome database	3,138	507	-	2,991	433	-	2,843	414	-	2,834	424	-
After exclusion of SNV/indel in 5'UTR, 3'UTR, downstream, upstream, intron and synonymous locations without local splice effect prediction	506	44	-	525	34	-	468	48	-	473	44	-
After exclusion of missense without SIFT, PPH2 or PhastCons prediction	420	44	-	450	34	-	383	48	-	394	44	-
After selection of variants consistent with recessive transmission (compound heterozygous, homozygous variants)	1 homozygous variant (in <i>DGKZ</i> ) 0 heterozygous compound											

**Table S3.2. Summary of the whole genome sequencing results.**

SNV: Single nucleotide variation, indel: gain or loss of up to 50 nucleotides at a single locus, SV: Structural Variation. ROH: Region of homozygosity as defined by the SNParrays. Annotations are gathered using Alamut Batch v1.11, especially for the variation databases including gnomAD (v2.0.2, Oct. 2017), 1000 Genomes Project phase3 release (version 20150813 v5b) and the following predictions including phastCons (UCSC, 44 vertebrates). Effect on the splice has been evaluated using the MaxEntScan <sup>344</sup>, NNSPLICE 0.9 <sup>345</sup> and Splice Site Finder <sup>346</sup> by calculating score change between the wild type and the mutated sequences expressed as a percent differences. Missenses have been evaluated using default parameters from PolyPhen-2 (2.2.2) <sup>279</sup> and SIFT 4.0.3 <sup>347</sup>. Default cut-offs used have been described in VaRank <sup>348</sup> for both type of predictions. Exclusion of SV with a DGV (Gold standard from 20160515) frequency > 1 % is done only with studies of more than 1000 individuals.

	II.1			II.2			II.3			II.4			II.7		
	SNV	indel	SV	SNV	indel	SV	SNV	indel	SV	SNV	indel	SV	SNV	indel	SV
Total number of variants	4,013,210	1,147,432	4,376	4,018,208	1,136,593	10,695	3,966,388	1,097,328	8840	4,003,486	1,140,327	11,494	3,857,384	1,121,548	11,380
After selection of variant found in the ROH	3,426	1,250	6	2,531	1,101	10	3,048	1,148	6	2,646	1,065	9	2,635	1,111	6
After exclusion of variants not in the cataract and deafness candidate-genes list	816	304	3	652	312	4	723	292	2	666	287	2	663	283	2
After exclusion of variants with an allele frequency >1% (gnomAD, 1000G, internal exome database, DGV)	85	27	0	67	32	0	84	29	0	74	23	0	76	30	0
After exclusion of SNV/indel in 5'UTR, 3'UTR, downstream, upstream, intron and synonymous locations without local splice effect prediction	10	1	-	13	1	-	12	0	-	14	1	-	14	0	-
After selection of homozygous variants in the affected individuals and heterozygous or absent in the healthy individuals	6 homozygous variants in 5 genes ( <i>ATG13</i> , <i>CELF1</i> , <i>MADD</i> , <i>PSMC3</i> and <i>RPS6KB2</i> )														

**Table S3.3. Orthologous ID equivalent of *PSMC3*, *ACTG1*, *CHMP4B* and *GJB6* in human, mouse and yeast.**

For each species, official gene symbols are indicated with the Uniprot <sup>349</sup> entry name into parenthesis. HGNC: Hugo Gene Nomenclature Committee. \* To avoid confusion, we would like to highlight that in 2006 Binato *et al* <sup>350</sup> named the mouse orthologue of *PSMC3* “PRSA\_MOUSE” while it is currently named “PRS6A\_MOUSE”.

HGNC approved name	Human	Mouse	Yeast
Proteasome 26S subunit, ATPase 3	PSMC3 (PRS6A_HUMAN)	psmc3 (PRS6A_MOUSE*)	RPT5 (YOR117W)
Actin gamma 1	ACTG1 (ACTG_HUMAN)	actg1 (ACTG_MOUSE)	ACT1 (YFL039C)
Charged multivesicular body protein 4B	CHMP4B (CHM4B_HUMAN)	chmp4b (CHM4B_MOUSE)	SNF7 (YLR025W)
Gap junction protein beta 6	GJB6 (CXB6_HUMAN)	gjb6 (CXB6_MOUSE)	

**Table S3.4. Publications showing genes interactions between *PSMC3*, *ACTG1*, *CHMP4B* and *GJB6*.**

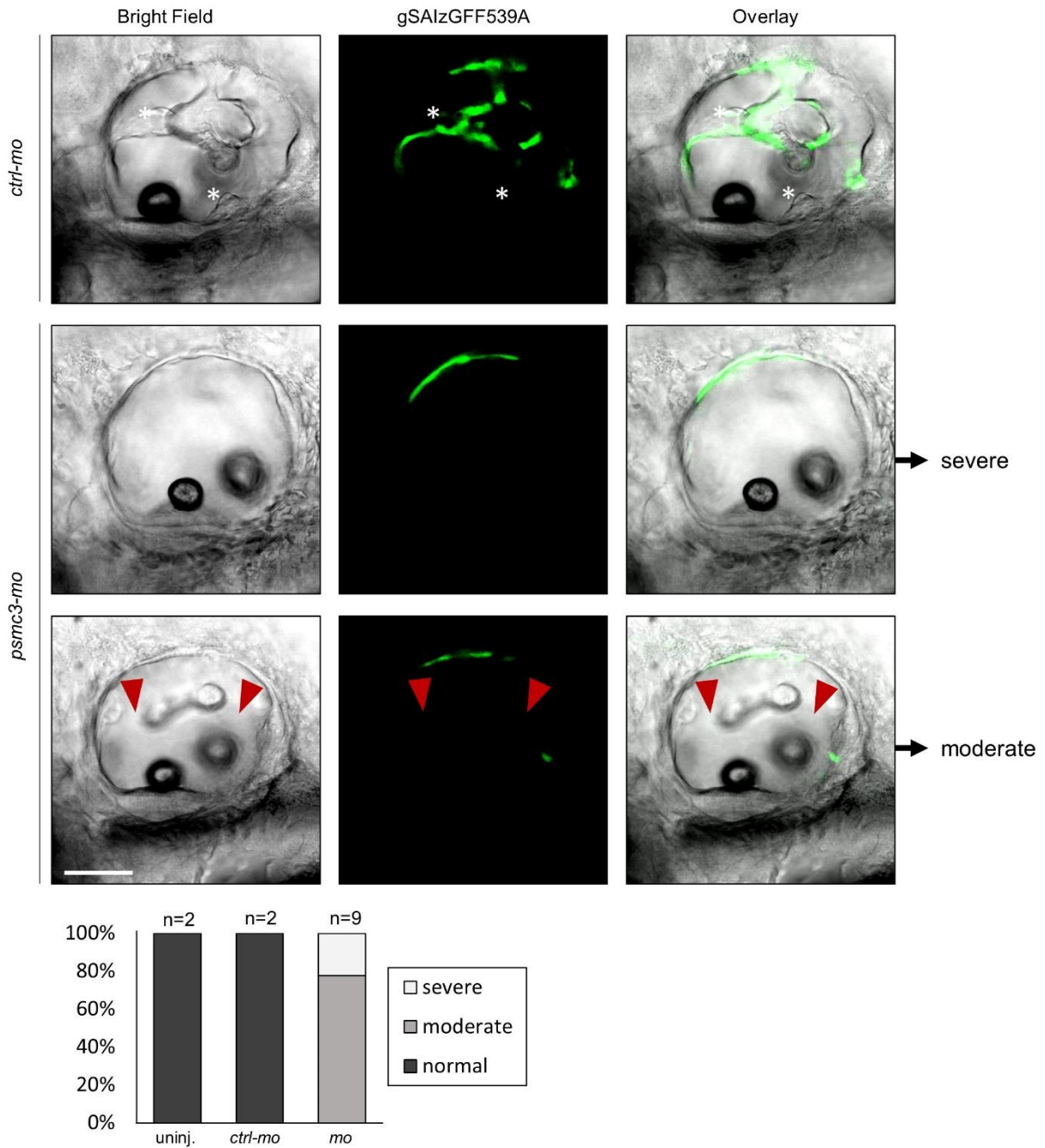
Human protein interaction	Publication showing the interaction	Organism of interest	Orthologous genes names used
PSMC3 / GJB6	Binato et al. <sup>350</sup>	Mouse	PRSA_MOUSE / CXB6_MOUSE
PSMC3 / ACTG1	Guerrero et al. <sup>351</sup>	Yeast	<i>RPT5</i> / <i>ACT1</i>
PSMC3 / CHMP4B	Wang et al. <sup>352</sup>	Yeast	YOR117W.1 / YLR025W.2

**Table S3.5. Proteasome proteins identified using nanoLC-MS/MS analysis and quantified by Spectral Count.**

PSMC3 was immunoprecipitated from control (CONTROL#1/2/3) and patients (HEALTHY#1/2/3 and AFFECTED#1/2/3) fibroblast protein lysates. Interaction partners were determined by searching against the complete UniProtKB Human proteome set <sup>349</sup>. The total number of MS/MS fragmentation spectra identified by Mascot algorithm <sup>353</sup> and validated at FDR<1 % were then counted for each protein in each sample. Spectral Count values were then normalized by the total number of spectra in each sample. in order to calculate the final relative abundance of each protein between the AFFECTED triplicates and the HEALTHY triplicates. H: Healthy, A: Affected.

UniProt Acc	Gene name	Description from UniProt	RAW Spectral Count									NORMALIZED Spectral Count						Mean #spectra		STDEV #spectra		Ratio (Mean)
			CONTROL			HEALTHY			AFFECTED			HEALTHY			AFFECTED			H	A	H	A	A/H
			#1	#2	#3	#1	#2	#3	#1	#2	#3	#1	#2	#3	#1	#2	#3					
PSA1_HUMAN	PSMA1	Proteasome subunit alpha type-1	0	0	0	41	44	49	64	61	57	44.02	43.29	48.06	63.49	60.78	55.83	45.1	60.0	2.6	3.9	1.33
PSA2_HUMAN	PSMA2	Proteasome subunit alpha type-2	0	0	0	27	28	25	33	31	32	28.99	27.55	24.52	32.73	30.89	31.35	27.0	31.7	2.3	1.0	1.17
PSA3_HUMAN	PSMA3	Proteasome subunit alpha type-3	0	0	0	28	27	27	43	41	44	30.06	26.56	26.48	42.65	40.85	43.10	27.7	42.2	2.0	1.2	1.52
PSA4_HUMAN	PSMA4	Proteasome subunit alpha type-4	0	0	0	31	31	35	51	47	46	33.28	30.50	34.33	50.59	46.83	45.06	32.7	47.5	2.0	2.8	1.45
PSA5_HUMAN	PSMA5	Proteasome subunit alpha type-5	0	0	0	14	18	16	24	20	20	15.03	17.71	15.69	23.81	19.93	19.59	16.1	21.1	1.4	2.3	1.31
PSA6_HUMAN	PSMA6	Proteasome subunit alpha type-6	0	0	0	32	37	43	49	56	56	34.36	36.40	42.17	48.61	55.80	54.85	37.6	53.1	4.1	3.9	1.41
PSA7_HUMAN	PSMA7	Proteasome subunit alpha type-7	0	0	0	33	40	42	60	54	49	35.43	39.36	41.19	59.52	53.81	48.00	38.7	53.8	2.9	5.8	1.39
PSB1_HUMAN	PSMB1	Proteasome subunit beta type-1	0	0	0	26	31	35	41	45	43	27.92	30.50	34.33	40.67	44.84	42.12	30.9	42.5	3.2	2.1	1.38
PSB2_HUMAN	PSMB2	Proteasome subunit beta type-2	0	0	0	13	14	15	31	31	25	13.96	13.77	14.71	30.75	30.89	24.49	14.1	28.7	0.5	3.7	2.03
PSB3_HUMAN	PSMB3	Proteasome subunit beta type-3	0	0	0	15	21	18	29	26	25	16.10	20.66	17.65	28.77	25.91	24.49	18.1	26.4	2.3	2.2	1.45
PSB4_HUMAN	PSMB4	Proteasome subunit beta type-4	0	0	0	13	12	12	30	28	24	13.96	11.81	11.77	29.76	27.90	23.51	12.5	27.1	1.3	3.2	2.16
PSB5_HUMAN	PSMB5	Proteasome subunit beta type-5	0	1	0	25	27	27	39	35	35	26.84	26.56	26.48	38.69	34.87	34.28	26.6	35.9	0.2	2.4	1.35
PSB6_HUMAN	PSMB6	Proteasome subunit beta type-6	0	0	0	6	8	9	17	16	11	6.44	7.87	8.83	16.86	15.94	10.77	7.7	14.5	1.2	3.3	1.88
PSB7_HUMAN	PSMB7	Proteasome subunit beta type-7	0	0	0	8	9	9	11	14	11	8.59	8.85	8.83	10.91	13.95	10.77	8.8	11.9	0.1	1.8	1.36
PSB8_HUMAN	PSMB8	Proteasome subunit beta type-8	0	0	0	10	12	14	21	19	15	10.74	11.81	13.73	20.83	18.93	14.69	12.1	18.2	1.5	3.1	1.50
PSB9_HUMAN	PSMB9	Proteasome subunit beta type-9	0	0	0	6	5	5	6	5	7	6.44	4.92	4.90	5.95	4.98	6.86	5.4	5.9	0.9	0.9	1.09
PSB10_HUMAN	PSMB10	Proteasome subunit beta type-10	0	0	0	4	6	6	4	5	6	4.29	5.90	5.88	3.97	4.98	5.88	5.4	4.9	0.9	1.0	0.92
PRS4_HUMAN	PSMC1	26S proteasome regulatory subunit 4	0	0	0	63	72	71	84	77	74	67.64	70.84	69.63	83.32	76.72	72.49	69.4	77.5	1.6	5.5	1.12
PRS7_HUMAN	PSMC2	26S proteasome regulatory subunit 7	0	0	0	76	94	84	112	108	102	81.60	92.49	82.38	111.10	107.61	99.91	85.5	106.2	6.1	5.7	1.24
PRS6A_HUMAN	PSMC3	26S proteasome regulatory subunit 6A	0	0	0	97	108	107	104	100	103	104.15	106.26	104.94	103.16	99.64	100.89	105.1	101.2	1.1	1.8	0.96
PRS6B_HUMAN	PSMC4	26S proteasome regulatory subunit 6B	0	0	0	85	86	99	103	96	104	91.26	84.61	97.09	102.17	95.66	101.87	91.0	99.9	6.2	3.7	1.10
PRS8_HUMAN	PSMC5	26S proteasome regulatory subunit 8	0	0	0	68	65	72	92	83	84	73.01	63.95	70.61	91.26	82.70	82.28	69.2	85.4	4.7	5.1	1.23
PRS10_HUMAN	PSMC6	26S proteasome regulatory subunit 10B	0	0	0	89	88	92	88	87	96	95.56	86.58	90.23	87.29	86.69	94.04	90.8	89.3	4.5	4.1	0.98
PSMD1_HUMAN	PSMD1	26S proteasome non-ATPase regulatory subunit 1	0	0	0	97	109	104	133	129	113	104.15	107.24	102.00	131.93	128.54	110.69	104.5	123.7	2.6	11.4	1.18
PSMD2_HUMAN	PSMD2	26S proteasome non-ATPase regulatory subunit 2	0	0	0	105	124	119	127	136	133	112.73	122.00	116.71	125.98	135.51	130.28	117.1	130.6	4.6	4.8	1.11
PSMD3_HUMAN	PSMD3	26S proteasome non-ATPase regulatory subunit 3	0	0	0	74	83	81	109	103	97	79.45	81.66	79.44	108.12	102.63	95.02	80.2	101.9	1.3	6.6	1.27
PSMD4_HUMAN	PSMD4	26S proteasome non-ATPase regulatory subunit 4	0	0	0	27	25	27	38	32	30	28.99	24.60	26.48	37.69	31.89	29.39	26.7	33.0	2.2	4.3	1.24
PSMD5_HUMAN	PSMD5	26S proteasome non-ATPase regulatory subunit 5	0	0	0	21	22	19	20	20	15	22.55	21.65	18.63	19.84	19.93	14.69	20.9	18.2	2.0	3.0	0.87
PSMD6_HUMAN	PSMD6	26S proteasome non-ATPase regulatory subunit 6	0	0	0	54	68	72	71	73	68	57.98	66.90	70.61	70.43	72.74	66.61	65.2	69.9	6.5	3.1	1.07
PSMD7_HUMAN	PSMD7	26S proteasome non-ATPase regulatory subunit 7	0	0	0	34	46	36	42	40	35	36.50	45.26	35.31	41.66	39.86	34.28	39.0	38.6	5.4	3.8	0.99
PSMD8_HUMAN	PSMD8	26S proteasome non-ATPase regulatory subunit 8	0	0	0	19	24	25	24	23	21	20.40	23.61	24.52	23.81	22.92	20.57	22.8	22.4	2.2	1.7	0.98
PSMD9_HUMAN	PSMD9	26S proteasome non-ATPase regulatory subunit 9	0	0	0	27	33	32	27	24	22	28.99	32.47	31.38	26.78	23.91	21.55	30.9	24.1	1.8	2.6	0.78
PSD10_HUMAN	PSMD10	26S proteasome non-ATPase regulatory subunit 10	0	0	0	4	5	5	4	2	2	4.29	4.92	4.90	3.97	1.99	1.96	4.7	2.6	0.4	1.2	0.56
PSD11_HUMAN	PSMD11	26S proteasome non-ATPase regulatory subunit 11	0	0	0	73	79	81	93	95	88	78.38	77.73	79.44	92.25	94.66	86.20	78.5	91.0	0.9	4.4	1.16
PSD12_HUMAN	PSMD12	26S proteasome non-ATPase regulatory subunit 12	0	0	0	54	62	62	62	66	57	57.98	61.00	60.81	61.50	65.76	55.83	59.9	61.0	1.7	5.0	1.02
PSD13_HUMAN	PSMD13	26S proteasome non-ATPase regulatory subunit 13	0	0	0	45	53	53	63	60	48	48.31	52.15	51.98	62.49	59.79	47.02	50.8	56.4	2.2	8.3	1.11
PSDE_HUMAN	PSMD14	26S proteasome non-ATPase regulatory subunit 14	0	0	0	35	35	36	34	34	32	37.58	34.44	35.31	33.73	33.88	31.35	35.8	33.0	1.6	1.4	0.92





**Movie S3. Life imaging of semicircular canal formation (56 and 72 hpf) in *ctrl-mo* and *mo* injected embryos.** Morpholino (MO)-mediated knock-down of Psmc3 resulted in abnormal ear development. Canal pillars form between 56-72 hpf in uninjected and control injected fish (*ctrl-mo*) but not in morphants (*mo*). White asterisks indicate fused pillars. Red arrowheads mark unfused projections. Severe = no outgrowth of projections of semicircular canals. Moderate = projections do not fuse. Scale bar = 50  $\mu$ m.





# General Discussion

## 5 General Discussion

In this thesis I report the identification of two novel candidate genes for neurosensory diseases and the confirmation of the implication of one gene in a neurodevelopmental disorder using Next-Generation Sequencing technologies. Moreover, the zebrafish served as a model organism to validate the link between the respective candidate genes and the phenotype described in human patients. In the following advantages and limitations of NGS technologies and zebrafish as a model for human mendelian diseases will be discussed.

### 5.1 Advantages and limitations researchers are facing using WES or WGS for genetic diagnosis

About 7,000 different monogenetic diseases have been described at the phenotypic level. The genetic cause of those diseases, however, was only identified in less than half of them. In particular, the rapid development of NGS technology in terms of speed, length of reads, throughput and cost reduction has led to the identification of many novel rare disease genes in the last decade.<sup>354</sup>

Since the whole exome comprises only about 1-2 % of the genome, but was estimated to contain 85 % of the disease-causing mutations, in the last years especially WES has been routinely used and shown to be a valuable technique for the identification of novel disease genes.<sup>121</sup>

Also, for the projects described in this thesis, we used WES in all three index families to identify the disease-causing variants. In two out of three cases this strategy was successful. In a boy with a severe syndromic form of intellectual disability, I was able to identify a disease-causing homozygous variant in the gene *PRMT9* (project 1). Moreover, a heterozygous variant in the gene *MYH10* was identified as potential disease-causing in three members of a family affected by a strong ocular phenotype associated with coloboma, ptosis, epicanthus inversus and hypertelorism (project 2). In project 3, however, no interesting candidate gene could be identified using WES.

In spite of all progress, WES also has its limitations. Approximately 2.4 % of the exons are not captured by the WES data, either due to the incomplete exon coverage of the capture kit or due to sequencing limitations, for example in GC rich regions.<sup>125</sup> But also inadequate coverage caused by poor sequencing, false negative variants caused by mismatched reads or a limited amount of variant calls due to complex indels can be causes why the mutated gene cannot be detected.<sup>125</sup> Moreover, structural variants and variants in intergenic/intragenic sequences cannot be either identified or with poor reliability (CNV).

To overcome such difficulties, WGS experiment can be used and has proven its utility in our laboratory, as e.g. for project 3 or for the identification of a frequent tandem duplication in the gene *IFT140* in individuals affected by a rare ciliopathy, the Mainzer-Salidino Syndrome.<sup>186</sup> In the past, especially in

consanguineous families WGS has successfully been used to provide a genetic diagnosis to families where previously WES has failed to identify the disease causing variant.<sup>355</sup> An advantage of WGS is the uniform coverage depth, allowing the detection of mutations also in intronic, intergenic and exonic regions that are often inadequately covered by WES. In project 3, the whole genome of three affected and two healthy individuals was sequenced and led to the identification of a deep intronic mutation in the gene *PSMC3*. Nevertheless, the identification and validation of the disease-causing gene *PSMC3* took more than 10 years. Whole Genome Sequencing is getting cheaper and the cost gap between WES and WGS is getting smaller and smaller. Therefore, the question came up if it wouldn't be better to directly sequence the whole genome.<sup>356</sup>

However, even if the cost per genome seems to decrease each year, a WGS experiment entails many hidden costs.<sup>357</sup> The sequencing of one single individual generates an enormous amount of data. About 3-4 million of variants can be found per genome. In comparison, in a WES experiment usually only 30,000-60,000 variants/ individual can be identified. The bioinformatic analysis, the confirmation by Sanger sequencing and further functional experiments as well as the storage of WGS data cause further hidden costs, which are not included in the sequencing costs.<sup>357</sup> A major challenge in the prioritization of potential disease-causing variants are, in particular, the recognition of thousands of variants without a pathogenic effect. Many mutation prediction programs are still in the early stages of their development, which still leads to misjudgments of the programs.<sup>358</sup> The enormous importance of deep intronic areas for the splicing process e.g. has only been reported recently.<sup>359</sup> In the coming years it will therefore be absolutely necessary to improve those tools.

Due to the high number of genetic variations identified by WES or WGS, strict filters are required and there is a risk of accidentally removing the disease-causing gene. In order to rule out this possibility and to ensure that the identified candidate gene is the sole trigger of a phenotype, it is essential to compare the phenotypes of affected patients of different families. A common problem when working with rare diseases, however, is often that laboratories work only with single families. To find additional individuals carrying mutations in the same gene of interest the exchange of clinical data and identified mutations between genetic counselors and researcher all over the world is required. Also, in this work, I faced the problem to work only with single families. A good way to find more individuals with mutations in the same gene are the online data bases Decipher and Genematcher. Decipher provides free access to phenotypic and/ or genomic data of 34903 patients and Genematcher helps to connect with researchers and clinicians who share the interest in the same gene. In all three projects, we used both tools in order to find additional families. How useful such tools can be we experienced in project 1. Thanks to Decipher and Genematcher we were able to find six further families with individuals carrying LoF mutations in the gene *PRMT9* and presenting a similar phenotype, which allowed us to confirm *PRMT9* as the disease-causing gene. In case of project 2 and 3, however, no further families could be

recruited. In those cases, the use of animal models, such as zebrafish, is required to prove a link between the candidate gene and the patient phenotype.

## 5.2 Zebrafish as a model organism to study human heredity diseases – advantages and limitations

In this thesis, the zebrafish model was used to examine three candidate genes for neurosensory and/or neurodevelopmental diseases. For this purpose, two different strategies were chosen: targeted gene knockdown *via* morpholinos and gene knockout *via* the CRISPR/Cas9 system.

Morpholinos efficiently bind their target mRNAs and either prevent their translation or affect splicing. In many publications, they have been successfully used to reproduce a patient's phenotype. For a long time, morpholinos were predominantly used to study the function of specific genes, then researchers started more and more to question the results obtained by morpholino knockdown assuming that they are leading to many off-target effects.<sup>360</sup> To verify the results obtained with morpholinos, stable knockout mutants were generated, e.g. using the CRISPR/Cas9 system. Also guide RNAs may bind to off-targets. By performing several outcrosses, however, the observation of side-effects in a later generation is rather unlikely. Surprisingly, however, the knockout of many genes did result in no or only very weak phenotype compared to the strong phenotype observed in knockdown models.<sup>360</sup> This phenomenon has been observed not only in zebrafish but also in other model organisms such as mouse, *Drosophila*, yeast, as well as in human cells.<sup>360</sup>

While performing both knockdown and knockout experiments I also observed in two out of three cases that the phenotype of morphants and mutants differed from each other. Several theories may explain this phenomenon. (1) In zygotic mutants, the phenotype might be rescued by the maternally provided wildtype mRNA, whereas morpholinos also block the maternal translation of the gene.<sup>361</sup> (2) Moreover, the phenotype might be caused by off-target effects of morpholino that lead to a phenotype that is not related to the knockdown of the target gene. (3) Also, genetic compensation consisting in upregulation of a similar or related gene taking over its function, could explain why many knockout mutants do not display a phenotype.<sup>360</sup> Trigger for the genetic compensation could be the mutated mRNA.<sup>360</sup> In previous zebrafish experiments it was shown, that nonsense mediated decay (NMD) might trigger genetic compensation. NMD is a protective mechanism of the cell and is intended to prevent the synthesis of faulty proteins and gain of function effects.<sup>362</sup> Interestingly, a study performed in zebrafish by El-Brolosy et al. showed that the genetic inactivation of the NMD factor *upf1* not only led to reduced mRNA decay but also resulted in a loss of transcriptional adaptation and thus to a more severe phenotype.<sup>363</sup>

In the first project, I used the zebrafish model to create a link between the dysfunction of *PRMT9* and Intellectual disability/autism. Moreover, I suspected a role of *PRMT9* during primary cilia biogenesis, as the cilia length in patients with *PRMT9* mutations differed from those of healthy controls. First, I generated a knockout model carrying a frameshift mutation leading to a premature stop codon. To exclude a maternal contribution, I examined maternal-zygotic mutants and found that *MZprmt9*<sup>-/-</sup> mutants did not display any obvious morphological anomalies. Some patients affected by *PRMT9* mutations displayed brain anomalies, whereas the brain formation of *MZprmt9*<sup>-/-</sup> mutants during early development was not affected. This result does not exclude the possibility that brain anomalies develop at later stages. Further experiments would be necessary in order to map the different brain areas of the adult fish to detect morphological abnormalities. However, a different behaviour of *MZprmt9*<sup>-/-</sup> mutants compared to wildtype fish indicated a potential link between *Prmt9* and autism. Typical characteristics of ciliopathy in zebrafish, such as a curved body axis or kidney cysts<sup>247</sup>, are not present in *MZprmt9*<sup>-/-</sup> mutants. Overall, the phenotype observed in *MZprmt9*<sup>-/-</sup> mutants is rather mild. Since NMD was observed in mutants, it might be the explanation for this is genetic compensation triggered through the degradation of mutant mRNA. To identify the genes that are able to partly take over the function, further experiments would be required to assess transcriptomic changes in *MZprmt9*<sup>-/-</sup> mutants.

In addition, a morpholino blocking the translation of *prmt9* was injected. Morpholinos have the advantage that they have not been reported to trigger genetic compensation. Interestingly, *prmt9* morphants displayed a slight delay during the brain formation compared to control injected embryos, as well as a curved body axis in about one quarter of them. This might be a first indication that *Prmt9* also affects the zebrafish biogenesis of the cilia. However, to exclude a potential off-target effect further experimentation would be necessary, e.g. by using a second morpholino or performing a rescue experiment (co-injection of the *prmt9* mRNA).

Similarly, in project 2, the CRISPR/Cas9 system as well as morpholinos were used to establish a link between the candidate gene and the ocular phenotype observed in affected patients. Using WES, a dominant missense mutation in the gene *MYH10* was identified in three family members affected by a Baraitser-Winter-like syndrome with coloboma, hypertelorism, ptosis, and epicanthus inversus. Unfortunately, no further families could be recruited with mutations in *MYH10* to prove that the identified mutation in the candidate gene is indeed the disease-causing factor.

Interestingly, in a previous study by Gutzmann and his colleagues, a morpholino was already used to examine the function of *myh10* in zebrafish. The authors reported that the reduced *myh10* expression led to an abnormal actin distribution in the midbrain-hindbrain-boundary, a curved body axis and abnormal eyes.<sup>282</sup> A coloboma phenotype is not described, as the ocular phenotype was not further investigated.

To further examine the eye anomalies caused by *myh10* dysfunction, I generated a *myh10* knockout mutant and used two different morpholinos to knockdown the *myh10* gene expression. In the CRISPR mutants' ocular phenotypes was neither observed in heterozygous *myh10*<sup>-/+</sup> fish nor in homozygous *myh10*<sup>-/-</sup> carriers. Even maternal-zygotic mutants did not present any obvious morphological anomalies, thus a rescue of the phenotype due to maternal contribution can be excluded. In zebrafish morphants, however, a curved body axis and eye abnormalities were observed as previously published by Gutzmann. But, although *myh10* morphants displayed smaller eyes and the closure of the optic fissure proceeded more slowly compared to non-injected siblings (with one day delay), no distinct coloboma phenotype was observed. Molecular differences between the zebrafish eye and the human eye might be explained by their independent evolution lasting about 400 million years, the lateral *versus* forward oriented position of the eyes and their different functions. Unlike humans, zebrafish vision is especially important for hunting food or escaping from predators.<sup>364</sup> Therefore, the human eye and the zebrafish eye have probably developed different mechanisms to adapt to their specific needs.

In addition, morphants presented an abnormal muscle phenotype with strongly unorganized muscle fibers, a phenotype that was not reported in human patients. In zebrafish, Myh10 has already been suggested previously to be an important component of the pre-myofibrilla in zebrafish.<sup>272</sup> Unlike humans, Myh10 seems to be important in the early development of zebrafish skeletal muscles.

Again, in project 2, the phenotype of morphants and mutants differed completely from each other. One reason for this discrepancy might be, that the morpholinos are causing off-target effects. On the other hand, the injection of two different morpholinos used in this study (start site and splice-site morpholino), plus the already published morpholino, lead to a similar phenotype. This supports the hypothesis that they are indeed specific. However, the best way to exclude off-target effect is to perform a rescue experiment by co-injecting a morpholino together with wildtype mRNA. Unfortunately, this was not possible due to the huge size of the mRNA (7446 bp). The fact that mutants did not present any obvious phenotype might be explained by other genes that are taking over the genes function. An argument for this theory is that I detected NMD in *MZmyh10*<sup>-/-</sup> mutants, which is a trigger for genetic compensation.

Project 3 is a good example, that the CRISPR/Cas9 system can be also successfully used to reproduce a phenotype seen in morphants. The goal of this project was to establish a link between *PSMC3* and the development of the eye and ear, as human patients with a deep intronic mutation in *PSMC3* displayed congenital deafness and pre-onset cataract. A morpholino and two different guide RNAs were used to decrease the gene expression of *psmc3* in zebrafish embryos. Due to the high efficiency

of the two guide RNAs of 54-55%, I could use the injected embryos (F0, crispants) for further investigations. In F0, the embryos are mosaic and cells contain different kinds of *psmc3* mutations. Both crispants and morphants recapitulate the cataract phenotype described in humans. In its structure and function, the zebrafish ear resembles the human inner ear. One major difference, however, is that the cochlea, the most important structure of human hearing, is absent in the zebrafish ear.<sup>135</sup> The cochlea is also the structure of the ear that is affected in the patient. As anomalies in the cochlea are responsible for the deafness phenotype in the patients, the zebrafish may not be the optimal model organism in this case. Nonetheless, morphological changes were observed in the zebrafish ear of *psmc3* morphants and crispants. *psmc3* deficiency seemed to have an impact on the semicircular canals, as they failed to fuse after 4 dpf. In addition, the number of kinocilia in the cristae was significantly reduced. Why a gene playing a major role in the hearing process in humans has an influence on the vestibular system in fish, is a question that cannot be answered yet. Interestingly, however, the dysfunction of the gene *dfna5*, the orthologue of a human deafness gene, also leads to unfused semicircular canals in zebrafish.<sup>146</sup> For this reason, further investigations of the signaling pathways in the zebrafish ear are of great interest.

Overall, in this study, the zebrafish model has been successfully used to establish a link between our three candidate genes and neurodevelopmental and neurosensory diseases. Using zebrafish as an animal model, however, can also lead to difficulties. A major disadvantage of morpholinos is that they may also bind to off-target effects. Therefore, good controls are necessary in order to rule out that the observed phenotype is a side effect, like using different morpholinos and performing rescue experiment. Nonetheless, it was recommended not to only rely on morpholinos. Morpholinos should rather be used as a control for stable mutants.<sup>361</sup> To avoid nonsense-mediated decay and thus genetic compensation in KO models, missense mutations or in-frame indels can be created in important regulatory domains or promotor regions can be deleted. However, even if the knockout of a gene does not lead to the expected phenotype, it is of high interest to find out which genes are able to compensate for the loss of a specific candidate gene, especially for the development of new therapeutics.

# VI

## Conclusion and Perspectives



## 6 Conclusion and Perspectives

Overall, I contributed to the identification of two new disease genes associated with neurosensory diseases and confirmed the implication of a gene in a neurodevelopmental disorder.

The results gained in this thesis can be an important step towards a better understanding of the pathophysiology of neurodevelopmental and neurosensory diseases. The better we understand the underlying mechanisms of a disease, the easier it will be to develop therapeutic strategies. Therefore, any new gene that can be assigned to a disease is a step in the right direction.

For the patients and relatives, the identification of the genetic cause often means a great relief, as families of rare genetic diseases often undergo a diagnostic odyssey over years. Even if there are no treatment options yet, the diagnosis often helps them to get a better understanding of the disease. They can get in touch with other affected families, often a prognosis is possible and helps the families to adapt emotionally to the future and parents with a desire to have another child can be better advised.

In the coming years, especially the use of WGS will be of great interest, as 50-75 % of patients affected by a Mendelian disorder did not receive a genetic diagnosis by WES.<sup>365</sup> To simplify the complex analysis of WGS data, it has been shown that RNA sequencing in parallel can be a very helpful tool to identify the causative gene mutation. This strategy especially simplifies the recognition of pathologically relevant mutations in noncoding sequences since the effect of the mutation on transcription can be directly observed.<sup>365</sup>

Although there's still a need of improving the data analysis and storage of NGS data especially of WGS data, there's no doubt that NGS approaches will play an important role in the future.

One of the main goals for the future will be to use WGS data for personalized medicine to find out which medications are best for a specific patient and which therapeutic approaches may not lead to an improvement of the patient's conditions.

Also, the zebrafish model is of great interest for future personalized medicine. As the zebrafish has been shown in many cases to be able to recapitulate a patient's phenotype, the model is suitable for therapeutic studies. To find the most effective therapy for an individual patient, therapies can first be tested on a zebrafish model that has been previously genetically manipulated and thus present the same mutations. In particular, in more complex diseases, such as cancer, for example, this strategy might be time-saving and would allow patients to benefit more quickly from an effective therapy.

Although it will take a few more years to actually understand the relationship of the genome to diseases, personalized medicine will definitely come. And every new gene that has been assigned to a disease is a small step in the right direction.

# VII

Résumé en français

## 7 Résumé en français

Ce résumé français résume brièvement le contexte scientifique, l'objectif de la thèse, la stratégie expérimentale et les résultats obtenus.

### Introduction

Grâce aux approches de séquençage de la prochaine génération (NGS : *Next Generation Sequencing*), d'innombrables nouveaux gènes de maladie ont été identifiés au cours de la dernière décennie. L'identification de ces gènes est essentielle pour comprendre le rôle physiologique d'un gène ou d'une protéine, pour mieux connaître la physiopathologie d'une maladie, pour trouver de nouvelles cibles thérapeutiques et pour fournir des conseils génétiques aux familles des parents touchés. Bien que le chemin soit très long entre l'identification d'un nouveau gène de maladie et le développement de thérapies appropriées, de sorte que les familles ne pourraient pas profiter d'une thérapie avant des décennies, avec l'identification du déclencheur génétique, les familles peuvent mieux évaluer le niveau de risque pour une autre grossesse ou pour un autre membre de la famille. En outre, l'évolution de la maladie peut être mieux estimée et les familles peuvent participer à des groupes d'entraide, ce qui leur donne le sentiment d'être soutenues et de ne pas être seules.

L'objectif de cette « Cotutelle de Thèse » co-directée par Hélène Dollfus, responsable du Laboratoire de Génétique Médicale UMRS\_1112 de l'Université de Strasbourg et Uwe Strähle, responsable de l'Institut de Toxicologie et de Génétique de l'Institut de Technologie de Karlsruhe, était l'identification de nouveaux gènes associés aux maladies rares neurodéveloppementales ou neurosensorielles.

Au cours des dernières années, de nombreuses collaborations fructueuses entre les deux laboratoires ont permis l'identification de nouveaux gènes de maladies, en particulier pour un groupe de maladies appelées ciliopathies.<sup>1-4</sup> Le laboratoire de génétique médicale UMRS\_1112 travaille en étroite collaboration avec le service de génétique médicale et le centre de référence pour les affections rares en génétique ophtalmologique (CARGO) du CHU de Strasbourg ce qui permet un contact direct avec les patients et un recrutement *ad hoc* pour la recherche. De plus, des analyses bioinformatiques et des expériences sur les cellules des patients ont été réalisées dans le laboratoire UMRS\_1112 Strasbourg, tandis que des modèles de poisson-zèbre appropriés pour les gènes candidats respectifs ont été générés à l'ITG de Karlsruhe.

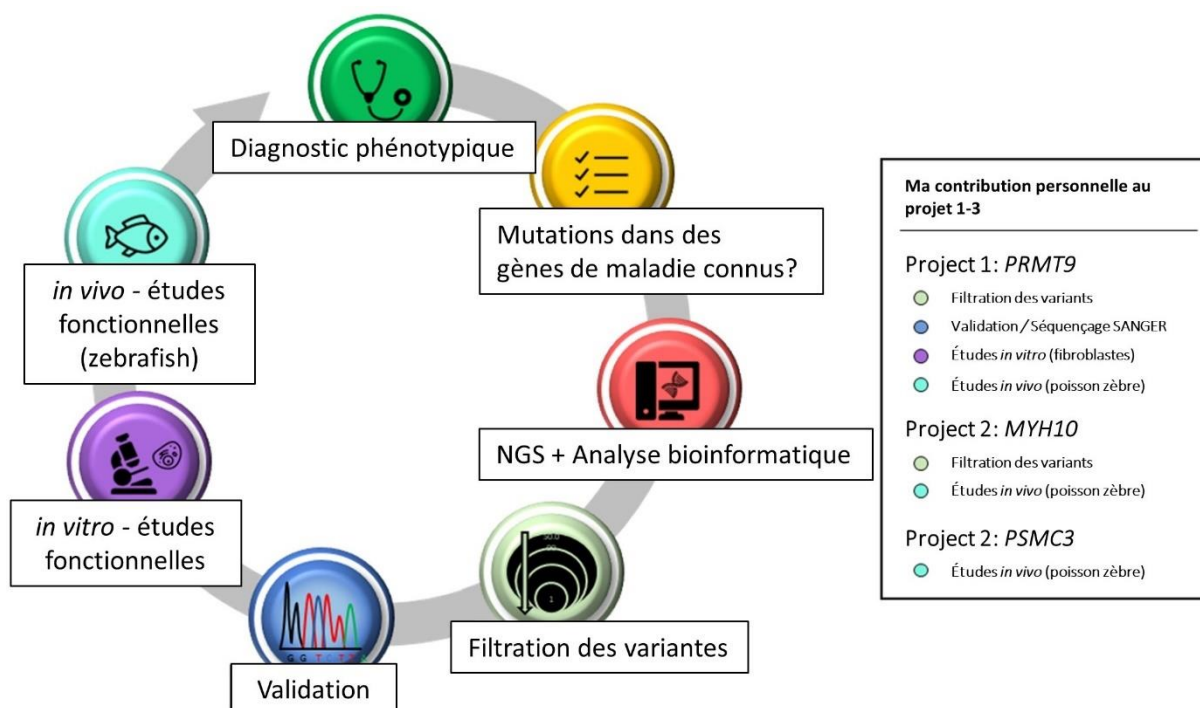
Ainsi, au cours de ma thèse de doctorat, j'ai eu l'occasion de bénéficier de l'expérience et de l'aide des deux laboratoires, ce qui m'a permis de contribuer à l'identification de trois nouveaux gènes de maladies associés à des maladies neurodéveloppementales ou neurosensorielles.

## Objectif de la thèse

Les maladies rares neurodéveloppementales et neurosensorielles peuvent affecter considérablement la qualité de vie des personnes touchées et de leurs proches, mais les options de traitement efficaces sont peu nombreuses. Il est donc très intéressant d'acquérir de nouvelles connaissances sur ces maladies, notamment pour tenter d'identifier des cibles et des voies thérapeutiques potentielles.

**L'objectif principal de cette thèse était, premièrement, de réaliser des investigations génétiques sur des patients atteints de maladies rares neurosensorielles ou neurodéveloppementales, afin d'identifier de nouveaux gènes de la maladie et, deuxièmement, de valider ces gènes candidats par des tests fonctionnels pour mieux comprendre la biologie de la maladie (expériences *in vitro* et *in vivo*).**

Dans cette thèse, je rapporte l'identification de trois gènes de maladies nouvelles associés à trois maladies neurosensorielles ou neurologiques ultra-rare. La stratégie que j'ai suivie pour identifier ces gènes et ma contribution personnelle aux trois projets distincts est illustrée dans la figure a.



**Figure a. Stratégie utilisée pour identifier les gènes de maladies et ma contribution personnelle aux trois projets de ma thèse.** Les icônes ont été créées à l'aide de l'icons8.com.

Pour mon étude, je n'ai considéré que les patients sans diagnose génétique et chez qui le séquençage de Sanger et le séquençage ciblé de la prochaine génération (NGS) effectués antérieurement étaient

négatifs. Pour identifier la mutation génétique causale, nous avons utilisé des approches NGS (WES ou/et WGS), suivies d'une analyse bioinformatique, d'un filtrage des variantes et de la validation des gènes candidats par séquençage Sanger. Des analyses fonctionnelles pour valider l'impact des variants identifiés et le rôle du gène ont été réalisés au sein du laboratoire de Génétique Médicale (Université de Strasbourg, UMRS 1112) en utilisant les fibroblastes de la peau des patients ainsi qu'en générant les modèles de *knock-down* ou *knockout* correspondants à l'Institut de Toxicologie et de Génétique (Karlsruher Institut für Technologie).

Grâce à cette stratégie, j'ai pu contribuer à l'identification de trois nouveaux gènes de maladie associés à des troubles neurodéveloppementaux et neurosensoriels :

- Partie 1. *PRMT9* -> Forme syndromique de la déficience intellectuelle
- Partie 2. *MYH10* -> Syndrome de Baraitser-Winter (phénotype oculaire fort avec colobome et ptose)
- Partie 3. *PSMC3* -> Syndrome neurosensoriel avec cataracte précoce et surdité congénitale

Les manuscrits de ces trois projets ont été ou vont être soumis dans les prochaines semaines et sont inclus dans cette thèse :

**(1) Ariane Kröll-Hermi et al. (2019) Biallelic variations in *PRMT9* delineate a novel syndromic form of intellectual.**

**[Manuscrit en préparation, sera soumis à l' « *American Journal of Human Genetics* »]**

**(2) Ariane Kröll-Hermi et al. (2019) *MYH10* dominant variant causes Baraitser-Winter cerebro-fronto-facial syndrome related phenotype with major ophthalmic developmental features.**

**[Manuscrit soumis au « *Journal of Medical Genetics* »]**

**(3) Ariane Kröll-Hermi et al. (2019) Proteasome subunit *PSMC3* variants cause neurosensory syndrome combining deafness and cataract due to proteotoxic stress.**

**[Manuscrit soumis au « *EMBO Molecular Medicine* »]**

## Résultats

### Projet 1

Les résultats du projet 1 sont également décrits dans l'article : « Biallelic variations in PRMT9 delineate a novel syndromic form of intellectual. » qui est actuellement en préparation et sera soumis au « American Journal of Human Genetics ».

La déficience intellectuelle (DI) sévère a une forte composante génétique.<sup>229</sup> Bien que plus de 700 gènes de DI aient déjà été identifiés dans les formes mendéliennes de DI isolées et syndromiques, les chercheurs estiment que ce nombre dépassera largement les 1000 dans les années à venir. Dans le projet 1, je présente un nouveau gène muté dans la déficience intellectuelle syndromique autosomique récessive.

Afin d'identifier la variation pathogène chez un patient algérien présentant une DI sévère, une hétérotopie neuronale périventriculaire bilatérale, une épilepsie, un trouble du spectre de l'autisme (TSA) et une polydactylie post-axiale bilatérale, un séquençage de l'exome entier a été réalisé. Une délétion d'un nucléotide (c.545delT) dans la séquence codante du gène *PRMT9* a été détectée, entraînant un décalage de cadre et un codon stop précoce (p.Leu182Trpfs\*) qui devrait donner naissance à une protéine sévèrement tronquée. En utilisant le séquençage de Sanger, il a été démontré que la mutation ségrége avec la maladie dans la famille. *PRMT9* appartient à une famille de protéines appelée les protéines arginine méthyltransférases (*PRMT*). À ce jour, neuf *PRMT* sont décrites et sont classées en fonction de leur produit de méthylation. Les *PRMT1*, *PRMT2*, *PRMT3*, *PRMT4/CARM1*, *PRMT6* et *PRMT8* sont des enzymes de type I et sont capables de monométyler (MMA) ou de dimétyler asymétriquement les résidus d'arginine (ADMA). *PRMT9* est, comme *PRMT5*, une enzyme de type II et produit des MMA et des diméthyl argénines symétriques (SDMA).<sup>167</sup> Les *PRMT* ont un large spectre de cibles, comme les histones, les protéines de liaison à l'ARN, les canaux ioniques, les protéines de transport, les protéines d'échafaudage, les facteurs de transcription, les coactivateurs et les protéines impliquées dans la transmission neuronale.<sup>230</sup> Ce jour, on ne connaît qu'un seul partenaire d'interaction de *PRMT9*, le facteur d'épissage *SAP145*, et on pense donc que *PRMT9* joue un rôle important dans l'épissage alternatif.<sup>3</sup> *PRMT9* a déjà été proposé comme gène candidat pour la DI isolé.<sup>170</sup> Cependant, un seul individu (aucune ségrégation familiale disponible à ce moment-là) et l'absence d'expériences fonctionnelles n'ont pas pu confirmer cette hypothèse.<sup>4</sup> En utilisant GeneMatcher et Decipher, j'ai pu recruter huit familles supplémentaires avec des patients porteurs de mutations dans le gène *PRMT9*.<sup>5,6</sup> Toutes présentaient une forme similaire de déficience intellectuelle syndromique. Outre la DI, les symptômes les plus fréquents étaient le retard global du développement, l'épilepsie, l'altération de l'articulation de la parole, l'autisme et l'hypotonie.

Comme le patient de référence présentait une polydactylie postaxiale bilatérale, des anomalies rénales, un retard de développement et était soupçonné d'être atteint d'une rétinite pigmentaire, il a d'abord été supposé d'être affecté par le syndrome de Bardet-Biedl (une ciliopathie). Les ciliopathies sont des pathologies associées à des anomalies du cil primaire, un organe cellulaire présent à la surface de presque toutes les cellules de l'organisme et jouant un rôle essentiel dans les voies de signalisation développementales et homéostatiques.<sup>231</sup> Afin d'examiner un effet potentiel d'un déficit de PRMT9 sur la ciliogenèse, les fibroblastes cutanés de trois patients et les cellules témoins ont été privés de sérum pour induire la biogenèse du cil primaire. En effet, les cils primaires chez les trois patients étaient significativement plus longs que dans les cellules témoins, ce qui suggère un rôle potentiel des PRMT9 pendant la ciliogenèse.

De plus, les premières expériences ont indiqué que des formes mutées de PRMT9 pourraient avoir un impact sur l'épissage alternatif. On sait que PRMT9 est responsable de la méthylation du facteur d'épissage SAP145, mais dans les expériences de Western Blot, la forme méthylée de SAP145 n'était détectable que dans les cellules témoins, ce qui montre que les formes mutées de PRMT9 de trois patients différents n'étaient pas capables de méthyler leur partenaire d'interaction.

Afin d'identifier les différences dans l'épissage alternatif et l'expression génique causées par les mutations de *PRMT9*, un séquençage de l'ARN a été effectué sur les cellules de trois patients et de trois témoins. Pour trouver un lien entre PRMT9 et les symptômes des patients, j'ai comparé la liste des gènes attribués aux bases de données sur les gènes de la déficience intellectuelle, de l'autisme et les gènes associés à la fonction et à la biogenèse des cils primaires. Des différences intéressantes dans l'expression des gènes ont ensuite été validées par PCR quantitatif. Par exemple, les gènes candidats pour la déficience intellectuelle et l'autisme, *PCDH10*, *SHROOM4* et *PLXDC2*, sont exprimés beaucoup plus fortement dans les cellules des patients que dans les fibroblastes témoins, ce qui renforce l'hypothèse selon laquelle *PRMT9* est un gène de la déficience intellectuelle. L'induction de la ciliogenèse par privation de sérum dans les cellules des patients par rapport aux cellules témoins a entraîné la dérégulation de l'expression de l'ARNm de gènes de cils bien connus tels que *GLI2* et *RPGRIP1L*, ce qui indique que *PRMT9* pourrait jouer un rôle pendant la biogenèse des cils primaires et pourrait avoir un impact sur la voie de signalisation des hérissés.

De plus, j'ai utilisé le système CRISPR/Cas9 pour générer un modèle de *knockout prmt9* de poisson-zèbre. Les mutants zygotiques maternels (*MZ*)*prmt9*<sup>-/-</sup> ne présentent pas de phénotype de ciliopathie et le développement général du cerveau ne semble pas être affecté. Cependant, après avoir effectué un test de comportement social chez les adultes *MZprmt9*<sup>-/-</sup>, la vitesse des mutants a été significativement réduite par rapport aux poissons témoins. De plus, contrairement aux animaux

témoins, les mutants ne semblaient pas s'agiter dans le voisinage immédiat de leurs congénères, ce qui pourrait indiquer un phénotype d'autisme.

En conclusion, dans ce premier article, je rapporte l'identification d'une nouvelle mutation pathogène dans le gène *PRMT9* qui provoque une forme syndromique de DI autosomique récessive. Cette nouvelle découverte peut être une étape importante vers une meilleure compréhension de la physiopathologie des troubles neurodéveloppementaux et pourrait éventuellement aider à la recherche de cibles et de voies thérapeutiques.

## Projet 2

Les résultats du projet 2 sont également décrits dans l'article : « *MYH10* dominant variant causes Baraitser-Winter cerebro-fronto-facial syndrome related phenotype with major ophthalmic developmental features » qui était soumis au « Journal of Medical Genetics ».

Dans le monde, seulement une personne sur 100 000 est atteinte d'un colobome.<sup>81</sup> Le colobome est habituellement causé par l'échec de la fermeture d'une fissure optique pendant la 5e à la 7e semaine de vie du fœtus.<sup>80</sup> La combinaison du colobome et des anomalies périorbitaires est encore plus rare et n'a été observée que dans très peu de syndromes, comme le syndrome cérébrofronto-facial de Baraitser Winter (BWCF).<sup>270</sup>

Une combinaison de colobome, d'hypertélorisme, de ptose et d'epicanthus inversus a également été observée dans une des familles sur lesquelles nous travaillions. À l'aide du WES, nous n'avons pas pu identifier une mutation dans l'un des gènes associés au BWCF, *ACTB* ou *ACTG1*, mais nous avons identifié une mutation hétérozygote de faux-sens dans le gène *MYH10* (p. [Arg1471Pro] ; [=]) chez les trois membres touchés de la famille. *MYH10* code pour une myosine non musculaire et est une protéine motrice dépendante de l'actine et, curieusement, est censée interagir avec l'*ACTB* et l'*ACTG1* (base de données humaine FunCoup 4.0).<sup>271</sup>

Afin d'étudier l'impact de la mutation de faux sens hétérozygote sur la fonction de la protéine, des Western Blots et des immunomarquages ont été réalisés, montrant un taux de protéine *MYH10* diminué dans les cellules du patient par rapport aux cellules témoins. De plus, les manipulations ont révélé que *MYH10* co-localise avec les fibres d'actine et que les fibres d'actine dans les cellules du patient sont plus courtes et désorganisées par rapport aux fibres d'actine dans les cellules témoins.

Pour mieux comprendre la fonction du gène et pour prouver le lien entre la déficience en *myh10* et le phénotype des patients, j'ai d'abord créé un mutant *knockout* de poisson zèbre pour *myh10*. Les mutants *myh10*<sup>-/-</sup> ne sont pas mortels, survivent jusqu'à l'âge adulte et deviennent sexuellement



matures. Étonnamment, aucune anomalie évidente n'a été observée. Les embryons de poisson zèbre *myh10*<sup>-/-</sup> se développent normalement et ne semblent pas présenter un phénotype oculaire suggérant une compensation génétique probablement induite par une dégradation des ARNm non-sens. Pour éviter la compensation génétique, une deuxième stratégie a été utilisée pour réduire l'expression de *myh10* : le *knockdown via* morpholino. Un morpholino bloquant la traduction de la protéine ainsi qu'un morpholino modifiant l'épissage de pré-ARNm ont été injectés dans des embryons au stade unicellulaire et ont donné dans les deux cas un phénotype musculaire fort avec des fibres musculaires non organisées et un axe corporel courbé. Un résultat très intéressant, car il a déjà été suggéré précédemment que Myh10 est un composant important de la pré-myofibrille chez le poisson-zèbre.<sup>272</sup> Myh10 semble donc être important dans le développement précoce du muscle squelettique du poisson-zèbre, très probablement en association avec l'actine. En outre, les morphants ont une taille d'œil plus petite et la fissure optique se referme avec un jour de retard. Un phénotype distinct de colobome n'a pas été observé.

En conclusion, dans le projet 2, nous rapportons l'identification d'une nouvelle mutation hétérozygote de faux sens identifiée dans le gène *MYH10*, qui entraîne un syndrome de type BWCF. De plus, nous montrons que le déficit en *myh10* provoque une longueur et une organisation anormales de l'actine dans les cellules du patient et affecte le développement de l'œil et du muscle chez le poisson-zèbre.

### Projet 3

Les résultats du projet 2 sont également décrits dans l'article : « Proteasome subunit *PSMC3* variants cause neurosensory syndrome combining deafness and cataract due to proteotoxic stress » qui a été soumis au journal « EMBO Molecular Medicine ».

La cataracte précoce et la surdité congénitale sont une combinaison de symptômes très rare, qui n'a été observée que dans très peu de syndromes, par exemple le syndrome Aymé-Grip et le syndrome Wolfram.<sup>300,301</sup> Dans le projet 3, je fais état d'une grande famille consanguine d'origine turque avec trois enfants atteints de cataracte précoce, de surdité congénitale et d'autisme. Pour identifier la mutation du gène pathogène, on a effectué un test d'évaluation de l'environnement chez trois patients, mais on n'a pas réussi à révéler un gène candidat qui pourrait causer la combinaison de symptômes observée. Par conséquent, une EEM a été effectuée chez tous les patients atteints et chez deux personnes saines, ce qui a permis d'identifier une mutation intronique profonde homozygote dans le gène *PSMC3* (c.1127+337A>G), qui devrait créer un nouveau site donneur d'épissage.

Le gène *PSMC3* code pour la sous-unité régulatrice de protéase 26S 6A (Rtp5) et est l'une des six sous-unités connues du protéasome 26 AAA-ATPase formant l'anneau ATPase (Rpt1-6). En tant que partie du système du protéasome de l'ubiquitine (UPS), *PSMC3* est impliqué dans la régulation des niveaux de protéines intracellulaires et le contrôle de la qualité des protéines et joue donc un rôle important dans l'homéostasie cellulaire.<sup>43</sup>

Ce jour, des dysfonctionnements de l'UPS ont été particulièrement observés dans les maladies neurodégénératives, telles que la maladie d'Alzheimer, la maladie de Huntington (MIM : 143100), la maladie de Parkinson (MIM : 168600) ou la sclérose latérale amyotrophique (MIM : 105400).

Afin d'étudier l'effet de la mutation intronique profonde dans le gène *PSMC3* trouvée chez les trois individus affectés sur l'expression de l'ARNm, une RT-PCR a été réalisée et a confirmé la présence de la forme d'ARNm naturelle ainsi qu'une forme épissée alternative dans les cellules du patient. L'épissage alternatif est causé par la mutation intronique profonde qui conduit à l'inclusion d'un exon cryptique de la taille de 114 pb et qui résulte donc en un codon stop prématuré. De plus, une PCR quantitative en temps réel a été réalisée et a révélé une régulation négative de l'ARNm de *PSMC3* de type sauvage dans les cellules des patients ainsi que la présence d'une forme tronquée supplémentaire. Cependant, le niveau protéique et la localisation de *PSMC3* ne semblaient pas affectés par la mutation intronique profonde. La protéine tronquée semble être instable, car elle n'était pas détectable par Western Blot.

Compte tenu de l'importance de *PSMC3* pour la dégradation des protéines ubiquitaires, l'impact de la mutation intronique profonde sur la fonction de l'UPS a été étudié. Il est intéressant de noter que dans des conditions normales, les fibroblastes dérivés de patients porteurs de la mutation *PSMC3* homozygote présentaient une quantité accrue de complexes protéasomes. La fonction du protéasome ne semblait pas être affectée. Cependant, dans des conditions de stress, les fibroblastes des patients présentent une voie de traitement TCF11/Nrf1 épuisée qui les empêche de réguler à la hausse les sous-unités du protéasome et donc de préserver l'homéostasie protéique en raison de la déficience du protéasome.

De plus, le poisson zèbre, un modèle animal connu pour les maladies oculaires et la surdité chez l'homme, a été utilisé pour étudier le lien entre la cataracte/surdité et la déficience de *PSMC3*. Le gène du poisson-zèbre *psmc3* a une identité de séquence élevée de 83 % avec l'orthologue humain. Pour générer un modèle de *knockdown psmc3*, j'ai utilisé un oligonucléotide antisens morpholino ciblant un site d'épissage dans *psmc3*. Après quatre jours post fertilisation, le développement de la lentille et de l'oreille du poisson zèbre a été évalué chez des embryons vivants. Chez les morphants *psmc3*, un fort phénotype de cataracte, ainsi que des anomalies de fusion du canal semi-circulaire dans l'oreille du poisson zèbre, ont été observés. En revanche, les embryons injectés avec un morpholino témoin (avec

un échange de 5 paires de bases) se sont développés normalement. Pour évaluer les traits autistiques, j'ai effectué une hybridation *in situ* de montures entières en utilisant des marqueurs cérébraux, mais je n'ai pas pu observer de différence évidente entre les morpholinos et un groupe témoin. Pour confirmer ces résultats, j'ai utilisé la technique CRISPR/Cas9 pour éditer le génome du poisson zèbre. La génération F0 est souvent génétiquement mosaïque, cependant, dans le cas des ARN guides hautement efficaces, il a été démontré que les F0s, aussi appelés crispants, peuvent récapituler avec succès les phénotypes mutants.<sup>8-10</sup> De plus, dans ce projet, les crispants ont récapitulé le phénotype de l'œil et de l'oreille observé dans les morphants *psmc3*. Une expérience de sauvetage a confirmé la spécificité des morpholino et ARN guide, car la co-injection d'ARN messenger *psmc3* pouvait sauver le phénotype de l'œil et de l'oreille chez les morphants et les crispants.

Dans l'ensemble, nous concluons que la déficience en *PSMC3* est associée au développement de la cataracte et de la surdit  et sugg rons un r le important de *PSMC3* au cours du d veloppement du cristallin et de l'oreille chez l'homme et le poisson z bre.

## Conclusion

Dans l'ensemble, j'ai contribué à l'identification de deux nouveaux gènes associés à des maladies neurosensorielles et j'ai confirmé l'implication d'un gène dans un trouble du développement neurologique.

Les résultats obtenus dans le cadre de cette thèse peuvent constituer une étape importante vers une meilleure compréhension de la physiopathologie des maladies neurodéveloppementales et neurosensorielles. Le mieux nous comprenons les mécanismes sous-jacents d'une maladie, plus il sera facile de développer des stratégies thérapeutiques. Par conséquent, tout nouveau gène qui peut être attribué à une maladie est un pas dans la bonne direction.

Pour les patients et leurs proches, l'identification de la cause génétique est souvent un grand soulagement, car les familles de maladies génétiques rares subissent souvent une odyssée diagnostique pendant des années. Même s'il n'existe pas encore d'options thérapeutiques, la diagnose les aide à mieux comprendre la maladie. Ils peuvent entrer en contact avec d'autres familles touchées, souvent un pronostic est possible et aide les familles à s'adapter émotionnellement à l'avenir et les parents qui souhaitent avoir un autre enfant peuvent être mieux conseillés.

Dans les années à venir, l'utilisation du WGS sera particulièrement intéressante, car 50 à 75 % des patients atteints d'une maladie mendélienne n'ont pas reçu de diagnostic génétique par WES.<sup>365</sup> Pour simplifier l'analyse complexe des données du WGS, il a été démontré que le séquençage parallèle de l'ARN peut être un outil très utile pour identifier la mutation génétique causale. Cette stratégie simplifie particulièrement la reconnaissance des mutations pathologiquement pertinentes dans les séquences non codantes puisque l'effet de la mutation sur la transcription peut être directement observé.<sup>365</sup>

Bien qu'il soit encore nécessaire d'améliorer l'analyse et le stockage des données NGS, en particulier des données des WGS, il n'y a aucun doute que les approches des NGS joueront un rôle important à l'avenir.

L'un des principaux objectifs pour l'avenir sera d'utiliser les données WGS pour la médecine personnalisée afin de déterminer quels médicaments sont les meilleurs pour un patient donné et quelles approches thérapeutiques ne sont pas capable d'améliorer l'état du patient.

De plus, le modèle du poisson zèbre est d'un grand intérêt pour la future médecine personnalisée.

Comme il a été démontré dans de nombreux cas que le poisson-zèbre est capable de récapituler le phénotype d'un patient, le modèle est adapté aux études thérapeutiques. Pour trouver la thérapie la plus efficace pour un patient individuel, les thérapies peuvent d'abord être testées sur un modèle de

poisson-zèbre qui a été préalablement manipulé génétiquement et qui présente donc les mêmes mutations. En particulier, dans le cas de maladies plus complexes, comme le cancer par exemple, cette stratégie pourrait faire gagner du temps et permettrait aux patients de bénéficier plus rapidement d'une thérapie efficace.

Même s'il faudra encore quelques années pour comprendre réellement la relation entre le génome et les maladies, la médecine personnalisée viendra certainement. Et chaque nouveau gène qui est attribué à une maladie signifie un petit pas dans la bonne direction.

**VIII**

**References**

## 8 References

1. Stoetzel, C. *et al.* A mutation in VPS15 (PIK3R4) causes a ciliopathy and affects IFT20 release from the cis-Golgi. *Nature Communications* **7**, (2016).
2. Scheidecker, S. *et al.* Exome sequencing of Bardet–Biedl syndrome patient identifies a null mutation in the BBSome subunit *BBIP1* (*BBS18*). *Journal of Medical Genetics* **51**, 132–136 (2014).
3. Bloch-Zupan, A. *et al.* Homozygosity Mapping and Candidate Prioritization Identify Mutations, Missed by Whole-Exome Sequencing, in *SMOC2*, Causing Major Dental Developmental Defects. *The American Journal of Human Genetics* **89**, 773–781 (2011).
4. Scheidecker, S. *et al.* Mutations in *TUBGCP4* Alter Microtubule Organization via the  $\gamma$ -Tubulin Ring Complex in Autosomal-Recessive Microcephaly with Chorioretinopathy. *The American Journal of Human Genetics* **96**, 666–674 (2015).
5. Vaz, R., Hofmeister, W. & Lindstrand, A. Zebrafish Models of Neurodevelopmental Disorders: Limitations and Benefits of Current Tools and Techniques. *International Journal of Molecular Sciences* **20**, 1296 (2019).
6. Tărlungeanu, D. C. & Novarino, G. Genomics in neurodevelopmental disorders: an avenue to personalized medicine. *Experimental & Molecular Medicine* **50**, (2018).
7. Parsamanesh, N. & Miri-Moghaddam, E. Novel Insight Into Intellectual Disability; A Review Article. *Gene, Cell and Tissue In Press*, (2018).
8. Einfeld, S. L., Ellis, L. A. & Emerson, E. Comorbidity of intellectual disability and mental disorder in children and adolescents: A systematic review. *Journal of Intellectual & Developmental Disability* **36**, 137–143 (2011).
9. Johns, P. Dementia. in *Clinical Neuroscience* 145–162 (Elsevier, 2014). doi:10.1016/B978-0-443-10321-6.00012-6.
10. Nemerimana, M., Chege, M. N. & Odhiambo, E. A. Risk Factors Associated with Severity of Nongenetic Intellectual Disability (Mental Retardation) among Children Aged 2–18 Years Attending Kenyatta National Hospital. *Neurology Research International* **2018**, 1–11 (2018).
11. Milani, D., Ronzoni, L. & Esposito, S. Genetic Advances in Intellectual Disability. *Journal of Pediatric Genetics* **04**, 125–127 (2015).
12. Robertson, J., Hatton, C., Emerson, E. & Baines, S. Prevalence of epilepsy among people with intellectual disabilities: A systematic review. *Seizure* **29**, 46–62 (2015).
13. Harris, J. C. *Intellectual disability: a guide for families and professionals*. (2010).
14. Srivastava, A. K. & Schwartz, C. E. Intellectual disability and autism spectrum disorders: Causal genes and molecular mechanisms. *Neuroscience & Biobehavioral Reviews* **46**, 161–174 (2014).
15. Kalkan, R. The Use of Molecular Cytogenetic Techniques for the Identification of Chromosomal Abnormalities. in *Chromosomal Abnormalities - A Hallmark Manifestation of Genomic Instability* (eds. Larramendy, M. L. & Soloneski, S.) (InTech, 2017). doi:10.5772/67415.
16. Rauch, A. *et al.* Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *American Journal of Medical Genetics Part A* **140A**, 2063–2074 (2006).
17. Regan, R. & Willatt, L. Mental Retardation: Definition, Classification and Etiology. in *Monographs in Human Genetics* (ed. Knight, S. J. L.) vol. 18 16–30 (KARGER, 2010).
18. Tomac, V. *et al.* Etiology of Developmental Delay in the Pediatric Population. *Southeastern European Medical Journal* **1**, (2017).
19. Pérez Jurado, L. A., Peoples, R., Kaplan, P., Hamel, B. C. & Francke, U. Molecular definition of the chromosome 7 deletion in Williams syndrome and parent-of-origin effects on growth. *Am. J. Hum. Genet.* **59**, 781–792 (1996).
20. Vissers, L. E. L. M., Gilissen, C. & Veltman, J. A. Genetic studies in intellectual disability and related disorders. *Nature Reviews Genetics* **17**, 9–18 (2016).

21. Kochinke, K. *et al.* Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. *The American Journal of Human Genetics* **98**, 149–164 (2016).
22. Piton, A., Redin, C. & Mandel, J.-L. XLID-Causing Mutations and Associated Genes Challenged in Light of Data From Large-Scale Human Exome Sequencing. *The American Journal of Human Genetics* **93**, 368–383 (2013).
23. Verkerk, A. J. *et al.* Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905–914 (1991).
24. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**, 223–228 (2015).
25. Thompson, A. J. *et al.* Brain MRI changes in phenylketonuria. Associations with dietary status. *Brain* **116 ( Pt 4)**, 811–821 (1993).
26. Wiczorek, D. Autosomal dominant intellectual disability. *medizinische genetik* **30**, 318–322 (2018).
27. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. Arlington, VA, American Association. (2013).
28. Baio, J. *et al.* Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR. Surveillance Summaries* **67**, 1–23 (2018).
29. Karimi, P., Kamali, E., Mousavi, S. & Karahmadi, M. Environmental factors influencing the risk of autism. *Journal of Research in Medical Sciences* **22**, 27 (2017).
30. Ritvo, E. R., Freeman, B. J., Mason-Brothers, A., Mo, A. & Ritvo, A. M. Concordance for the syndrome of autism in 40 pairs of afflicted twins. *Am J Psychiatry* **142**, 74–77 (1985).
31. Risch, N. *et al.* A Genomic Screen of Autism: Evidence for a Multilocus Etiology. *The American Journal of Human Genetics* **65**, 493–507 (1999).
32. Bailey, A. *et al.* Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* **25**, 63–77 (1995).
33. Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nature Genetics* **46**, 881–885 (2014).
34. Klei, L. *et al.* Common genetic variants, acting additively, are a major source of risk for autism. *Molecular Autism* **3**, 9 (2012).
35. Satterstrom, F. K. *et al.* Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *bioRxiv* (2019) doi:10.1101/484113.
36. Dent, E. W., Merriam, E. B. & Hu, X. The dynamic cytoskeleton: backbone of dendritic spine plasticity. *Curr. Opin. Neurobiol.* **21**, 175–181 (2011).
37. Fett-Conte, A. C., Bossolani-Martins, A. L. & Rosan, D. B. A. Etiology of Autism the Complexity of Risk Factors in Autism Spectrum Disorder. in *Autism Spectrum Disorder - Recent Advances* (ed. Fitzgerald, M.) (InTech, 2015). doi:10.5772/59109.
38. Fassio, A. *et al.* SYN1 loss-of-function mutations in autism and partial epilepsy cause impaired synaptic function. *Hum. Mol. Genet.* **20**, 2297–2307 (2011).
39. Greco, B. *et al.* Autism-related behavioral abnormalities in synapsin knockout mice. *Behavioural Brain Research* **251**, 65–74 (2013).
40. Giannandrea, M. *et al.* Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly. *Am. J. Hum. Genet.* **86**, 185–195 (2010).
41. Milh, M. *et al.* Epileptic and nonepileptic features in patients with early onset epileptic encephalopathy and STXBP1 mutations. *Epilepsia* **52**, 1828–1834 (2011).
42. Piton, A. *et al.* Mutations in the calcium-related gene IL1RAPL1 are associated with autism. *Hum. Mol. Genet.* **17**, 3965–3974 (2008).
43. Zolk, O., Schenke, C. & Sarikas, A. The ubiquitin–proteasome system: Focus on the heart. *Cardiovascular Research* **70**, 410–421 (2006).



44. Hug, N., Longman, D. & Cáceres, J. F. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research* **44**, 1483–1495 (2016).
45. Addington, A. M. *et al.* A novel frameshift mutation in UPF3B identified in brothers affected with childhood onset schizophrenia and autism spectrum disorders. *Mol. Psychiatry* **16**, 238–239 (2011).
46. Glessner, J. T. *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**, 569–573 (2009).
47. Kleefstra, T., Schenck, A., Kramer, J. M. & van Bokhoven, H. The genetics of cognitive epigenetics. *Neuropharmacology* **80**, 83–94 (2014).
48. Kishi, N. & Macklis, J. D. MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. *Molecular and Cellular Neuroscience* **27**, 306–321 (2004).
49. Mattioli, F. *et al.* Mutations in Histone Acetylase Modifier BRPF1 Cause an Autosomal-Dominant Form of Intellectual Disability with Associated Ptosis. *Am. J. Hum. Genet.* **100**, 105–116 (2017).
50. Schoch, H. *et al.* Sociability Deficits and Altered Amygdala Circuits in Mice Lacking Pcdh10, an Autism Associated Gene. *Biological Psychiatry* **81**, 193–202 (2017).
51. Eltokhi, A., Rappold, G. & Sprengel, R. Distinct Phenotypes of Shank2 Mouse Models Reflect Neuropsychiatric Spectrum Disorders of Human Patients With SHANK2 Variants. *Frontiers in Molecular Neuroscience* **11**, (2018).
52. Wang, X. *et al.* Synaptic dysfunction and abnormal behaviors in mice lacking major isoforms of Shank3. *Human Molecular Genetics* **20**, 3093–3108 (2011).
53. Keeling, E., Lotery, A., Tumbarello, D. & Ratnayaka, J. Impaired Cargo Clearance in the Retinal Pigment Epithelium (RPE) Underlies Irreversible Blinding Diseases. *Cells* **7**, 16 (2018).
54. Kels, B. D., Grzybowski, A. & Grant-Kels, J. M. Human ocular anatomy. *Clinics in Dermatology* **33**, 140–146 (2015).
55. Sridhar, M. S. Anatomy of cornea and ocular surface. *Indian J Ophthalmol* **66**, 190–194 (2018).
56. Delamere, N. A. Ciliary Body and Ciliary Epithelium. *Adv Organ Biol* **10**, 127–148 (2005).
57. Ludwig, P. E., Motlagh, M. & Czyz, C. N. Physiology, Eye. in *StatPearls* (StatPearls Publishing, 2019).
58. Sung, C.-H. & Chuang, J.-Z. The cell biology of vision. *J. Cell Biol.* **190**, 953–963 (2010).
59. Singh, M. & Tyagi, S. C. Genes and genetics in eye diseases: a genomic medicine approach for investigating hereditary and inflammatory ocular disorders. *Int J Ophthalmol* **11**, 117–134 (2018).
60. Stoetzel, C. *et al.* BBS10 encodes a vertebrate-specific chaperonin-like protein and is a major BBS locus. *Nat. Genet.* **38**, 521–524 (2006).
61. Stoetzel, C. *et al.* Identification of a novel BBS gene (BBS12) highlights the major role of a vertebrate-specific branch of chaperonin-related proteins in Bardet-Biedl syndrome. *Am. J. Hum. Genet.* **80**, 1–11 (2007).
62. Leitch, C. C. *et al.* Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. *Nat. Genet.* **40**, 443–448 (2008).
63. Marion, V. *et al.* Exome sequencing identifies mutations in LZTFL1, a BBSome and smoothed trafficking regulator, in a family with Bardet-Biedl syndrome with situs inversus and insertional polydactyly. *J. Med. Genet.* **49**, 317–321 (2012).
64. Schaefer, E. *et al.* Mutations in SDCCAG8/NPHP10 Cause Bardet-Biedl Syndrome and Are Associated with Penetrant Renal Disease and Absent Polydactyly. *Mol Syndromol* **1**, 273–281 (2011).
65. Ferrari, S. *et al.* Retinitis pigmentosa: genes and disease mechanisms. *Curr. Genomics* **12**, 238–249 (2011).
66. Weihbrecht, K. *et al.* Keeping an Eye on Bardet-Biedl Syndrome: A Comprehensive Review of the Role of Bardet-Biedl Syndrome Genes in the Eye. *Med Res Arch* **5**, (2017).
67. Khan, L., Shaheen, N., Hanif, Q., Fahad, S. & Usman, M. Genetics of congenital cataract, its diagnosis and therapeutics. *Egyptian Journal of Basic and Applied Sciences* **5**, 252–257 (2018).

68. Santana, A. & Waiswo, M. The genetic and molecular basis of congenital cataract. *Arquivos Brasileiros de Oftalmologia* **74**, 136–142 (2011).
69. Sheeladevi, S., Lawrenson, J. G., Fielder, A. R. & Suttle, C. M. Global prevalence of childhood cataract: a systematic review. *Eye* **30**, 1160–1169 (2016).
70. Cvekl, A. & Zhang, X. Signaling and Gene Regulatory Networks in Mammalian Lens Development. *Trends in Genetics* **33**, 677–702 (2017).
71. Richardson, R. *et al.* Transcriptome profiling of zebrafish optic fissure fusion. *Scientific Reports* **9**, (2019).
72. McAvoy, J. W., Chamberlain, C. G., de longh, R. U., Hales, A. M. & Lovicu, F. J. Lens development. *Eye (Lond)* **13 ( Pt 3b)**, 425–437 (1999).
73. Bassnett, S. & Šikić, H. The lens growth process. *Progress in Retinal and Eye Research* **60**, 181–200 (2017).
74. Gill, D. *et al.* Genetic heterogeneity of the Coppock-like cataract: a mutation in CRYBB2 on chromosome 22q11.2. *Invest. Ophthalmol. Vis. Sci.* **41**, 159–165 (2000).
75. Graw, J. Genetics of crystallins: cataract and beyond. *Exp. Eye Res.* **88**, 173–189 (2009).
76. Retamal, M. A. *et al.* Connexin in Lens Physiology and Cataract Formation. *Journal of Clinical & Experimental Ophthalmology* **04**, (2013).
77. Kumari, S. S., Gandhi, J., Mustehsan, M. H., Eren, S. & Varadaraj, K. Functional characterization of an AQP0 missense mutation, R33C, that causes dominant congenital lens cataract, reveals impaired cell-to-cell adhesion. *Exp. Eye Res.* **116**, 371–385 (2013).
78. Zhang, L. *et al.* Progressive sutural cataract associated with a BFSP2 mutation in a Chinese family. *Mol. Vis.* **12**, 1626–1631 (2006).
79. Anand, D., Agrawal, S. A., Slavotinek, A. & Lachke, S. A. Mutation update of transcription factor genes FOXE3, HSF4, MAF, and PITX3 causing cataracts and other developmental ocular defects. *Hum. Mutat.* **39**, 471–494 (2018).
80. Nakamura, K. M., Diehl, N. N. & Mohny, B. G. Incidence, ocular findings, and systemic associations of ocular coloboma: a population-based study. *Arch. Ophthalmol.* **129**, 69–74 (2011).
81. Skalicky, S. E. *et al.* Microphthalmia, Anophthalmia, and Coloboma and Associated Ocular and Systemic Features: Understanding the Spectrum. *JAMA Ophthalmology* **131**, 1517 (2013).
82. Hornby, S. J., Adolph, S., Gilbert, C. E., Dandona, L. & Foster, A. Visual acuity in children with coloboma<sup>11</sup>The authors have no proprietary interest in any materials used in this study. *Ophthalmology* **107**, 511–520 (2000).
83. Huang, X.-F. *et al.* Unraveling the genetic cause of a consanguineous family with unilateral coloboma and retinoschisis: expanding the phenotypic variability of RAX mutations. *Scientific Reports* **7**, (2017).
84. Moosajee, M. & Gregory-Evans, C. Y. Advances in the molecular genetics of ocular coloboma. *Expert Review of Ophthalmology* **1**, 209–227 (2006).
85. Strömland, K. & Hellström, A. Fetal alcohol syndrome--an ophthalmological and socioeducational prospective study. *Pediatrics* **97**, 845–850 (1996).
86. Tellier, A. L. *et al.* Increased paternal age in CHARGE association. *Clin. Genet.* **50**, 548–550 (1996).
87. Wang, P., Liang, X., Yi, J. & Zhang, Q. Novel SOX2 mutation associated with ocular coloboma in a Chinese family. *Arch. Ophthalmol.* **126**, 709–713 (2008).
88. Deml, B. *et al.* Novel mutations in PAX6, OTX2 and NDP in anophthalmia, microphthalmia and coloboma. *Eur. J. Hum. Genet.* **24**, 535–541 (2016).
89. Bower, M. *et al.* Update of PAX2 mutations in renal coloboma syndrome and establishment of a locus-specific database. *Hum. Mutat.* **33**, 457–466 (2012).
90. Azuma, N. *et al.* Mutations of the PAX6 Gene Detected in Patients with a Variety of Optic-Nerve Malformations. *The American Journal of Human Genetics* **72**, 1565–1570 (2003).
91. Lalani, S. R. *et al.* Spectrum of CHD7 Mutations in 110 Individuals with CHARGE Syndrome and Genotype-Phenotype Correlation. *The American Journal of Human Genetics* **78**, 303–314 (2006).
92. Kelberman, D. *et al.* Mutation of SALL2 causes recessive ocular coloboma in humans and mice. *Human Molecular Genetics* **23**, 2511–2526 (2014).

93. Jamieson, R. V. Domain disruption and mutation of the bZIP transcription factor, MAF, associated with cataract, ocular anterior segment dysgenesis and coloboma. *Human Molecular Genetics* **11**, 33–42 (2002).
94. Schimmenti, L. A. *et al.* Novel mutation in sonic hedgehog in non-syndromic colobomatous microphthalmia. *Am. J. Med. Genet. A* **116A**, 215–221 (2003).
95. Asai-Coakwell, M. *et al.* GDF6, a Novel Locus for a Spectrum of Ocular Developmental Anomalies. *The American Journal of Human Genetics* **80**, 306–315 (2007).
96. Ye, M. *et al.* Mutation of the bone morphogenetic protein GDF3 causes ocular and skeletal anomalies. *Hum. Mol. Genet.* **19**, 287–298 (2010).
97. Deml, B. *et al.* Mutations in MAB21L2 Result in Ocular Coloboma, Microcornea and Cataracts. *PLOS Genetics* **11**, e1005002 (2015).
98. Casey, J. *et al.* First implication of STRA6 mutations in isolated anophthalmia, microphthalmia, and coloboma: A new dimension to the STRA6 phenotype. *Human Mutation* **32**, 1417–1426 (2011).
99. Roos, L. *et al.* A homozygous mutation in a consanguineous family consolidates the role of ALDH1A3 in autosomal recessive microphthalmia. *Clin. Genet.* **86**, 276–281 (2014).
100. Wang, L. *et al.* ABCB6 Mutations Cause Ocular Coloboma. *The American Journal of Human Genetics* **90**, 40–48 (2012).
101. Huang, X.-F. *et al.* Mutation of IPO13 causes recessive ocular coloboma, microphthalmia, and cataract. *Experimental & Molecular Medicine* (2018).
102. Williamson, K. A. *et al.* Heterozygous loss-of-function mutations in YAP1 cause both isolated and syndromic optic fissure closure defects. *Am. J. Hum. Genet.* **94**, 295–302 (2014).
103. Morton, C. C. & Nance, W. E. Newborn Hearing Screening — A Silent Revolution. *New England Journal of Medicine* **354**, 2151–2164 (2006).
104. Fortnum, H. M., Summerfield, A. Q., Marshall, D. H., Davis, A. C. & Bamford, J. M. Prevalence of permanent childhood hearing impairment in the United Kingdom and implications for universal neonatal hearing screening: questionnaire based ascertainment study. *BMJ* **323**, 536–540 (2001).
105. Watkin, P. & Baldwin, M. The longitudinal follow up of a universal neonatal hearing screen: the implications for confirming deafness in childhood. *Int J Audiol* **51**, 519–528 (2012).
106. Mathers, C., Smith, A. & Concha, M. Global burden of hearing loss in the year 2000.
107. Korver, A. M. H. *et al.* Congenital hearing loss. *Nature Reviews Disease Primers* **3**, (2017).
108. Kaga, K. Auditory nerve disease and auditory neuropathy spectrum disorders. *Auris Nasus Larynx* **43**, 10–20 (2016).
109. Marazita, M. L. *et al.* Genetic epidemiological studies of early-onset deafness in the U.S. school-age population. *Am. J. Med. Genet.* **46**, 486–491 (1993).
110. Mathur, P. & Yang, J. Usher syndrome: Hearing loss, retinal degeneration and associated abnormalities. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1852**, 406–420 (2015).
111. Zelante, L. *et al.* Connexin26 mutations associated with the most common form of non-syndromic neurosensory autosomal recessive deafness (DFNB1) in Mediterraneans. *Hum. Mol. Genet.* **6**, 1605–1609 (1997).
112. Riazuddin, S. *et al.* Tricellulin is a tight-junction protein necessary for hearing. *Am. J. Hum. Genet.* **79**, 1040–1051 (2006).
113. Wilcox, E. R. *et al.* Mutations in the gene encoding tight junction claudin-14 cause autosomal recessive deafness DFNB29. *Cell* **104**, 165–172 (2001).
114. Naz, S. *et al.* Mutations of ESPN cause autosomal recessive deafness and vestibular dysfunction. *J. Med. Genet.* **41**, 591–595 (2004).
115. Donaudy, F. *et al.* Espin gene (ESPN) mutations associated with autosomal dominant hearing loss cause defects in microvillar elongation or organisation. *J. Med. Genet.* **43**, 157–161 (2006).
116. Kitajiri, S. *et al.* Actin-bundling protein TRIOBP forms resilient rootlets of hair cell stereocilia essential for hearing. *Cell* **141**, 786–798 (2010).

117. Roux, I. *et al.* Otoferlin, defective in a human deafness form, is essential for exocytosis at the auditory ribbon synapse. *Cell* **127**, 277–289 (2006).
118. Collins, F. S. Positional cloning moves from perditional to traditional. *Nat. Genet.* **9**, 347–350 (1995).
119. Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Disease gene identification strategies for exome sequencing. *Eur. J. Hum. Genet.* **20**, 490–497 (2012).
120. Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **42**, 30–35 (2010).
121. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences* **106**, 19096–19101 (2009).
122. Lohmann, K. & Klein, C. Next Generation Sequencing and the Future of Genetic Diagnosis. *Neurotherapeutics* **11**, 699–707 (2014).
123. Biskup, S. & Gasser, T. Genetic testing in neurological diseases. *J. Neurol.* **259**, 1249–1254 (2012).
124. Metzker, M. L. Sequencing technologies — the next generation. *Nature Reviews Genetics* **11**, 31–46 (2010).
125. Li, M. H. *et al.* Utility and limitations of exome sequencing as a genetic diagnostic tool for conditions associated with pediatric sudden cardiac arrest/sudden cardiac death. *Hum. Genomics* **9**, 15 (2015).
126. Huckert, M. *et al.* Mutations in the latent TGF-beta binding protein 3 (LTBP3) gene cause brachyolmia with amelogenesis imperfecta. *Hum. Mol. Genet.* **24**, 3038–3049 (2015).
127. Foo, J. N., Liu, J. & Tan, E.-K. Next-generation sequencing diagnostics for neurological diseases/disorders: from a clinical perspective. *Hum. Genet.* **132**, 721–734 (2013).
128. Yu, Y., Wu, B. -I., Wu, J. & Shen, Y. Exome and Whole-Genome Sequencing as Clinical Tests: A Transformative Practice in Molecular Diagnostics. *Clinical Chemistry* **58**, 1507–1509 (2012).
129. Stark, Z. *et al.* Integrating Genomics into Healthcare: A Global Responsibility. *The American Journal of Human Genetics* **104**, 13–20 (2019).
130. Caspar, S. M. *et al.* Clinical sequencing: From raw data to diagnosis with lifetime value. *Clin. Genet.* **93**, 508–519 (2018).
131. Persani, L., de Filippis, T., Colombo, C. & Gentilini, D. GENETICS IN ENDOCRINOLOGY: Genetic diagnosis of endocrine diseases by NGS: novel scenarios and unpredictable results and risks. *European Journal of Endocrinology* **179**, R111–R123 (2018).
132. Streisinger, G., Walker, C., Dower, N., Knauber, D. & Singer, F. Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). *Nature* **291**, 293–296 (1981).
133. Bradford, Y. M. *et al.* Zebrafish Models of Human Disease: Gaining Insight into Human Disease at ZFIN. *ILAR Journal* **58**, 4–16 (2017).
134. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).
135. Blanco-Sánchez, B., Clément, A., Phillips, J. B. & Westerfield, M. Zebrafish models of human eye and inner ear diseases. in *Methods in Cell Biology* vol. 138 415–467 (Elsevier, 2017).
136. Gao, H. *et al.* Mecp2 regulates neural cell differentiation by suppressing the Id1 to Her2 axis in zebrafish. *J. Cell. Sci.* **128**, 2340–2350 (2015).
137. Golzio, C. *et al.* KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* **485**, 363–367 (2012).
138. Leblond, C. S. *et al.* Meta-analysis of SHANK Mutations in Autism Spectrum Disorders: a gradient of severity in cognitive impairments. *PLoS Genet.* **10**, e1004580 (2014).
139. Liu, C. *et al.* CRISPR/Cas9-induced shank3b mutant zebrafish display autism-like behaviors. *Molecular Autism* **9**, (2018).
140. Berryer, M. H. *et al.* Mutations in SYNGAP1 Cause Intellectual Disability, Autism, and a Specific Form of Epilepsy by Inducing Haploinsufficiency. *Human Mutation* **34**, 385–394 (2013).
141. Kozol, R. A. *et al.* Two knockdown models of the autism genes SYNGAP1 and SHANK3 in zebrafish produce similar behavioral phenotypes associated with embryonic disruptions of brain morphogenesis. *Human Molecular Genetics* **24**, 4006–4023 (2015).

142. Sakai, C., Ijaz, S. & Hoffman, E. J. Zebrafish Models of Neurodevelopmental Disorders: Past, Present, and Future. *Frontiers in Molecular Neuroscience* **11**, (2018).
143. Guo, S. Using zebrafish to assess the impact of drugs on neural development and function. *Expert Opinion on Drug Discovery* **4**, 715–726 (2009).
144. Dreosti, E., Lopes, G., Kampff, A. R. & Wilson, S. W. Development of social behavior in young zebrafish. *Frontiers in Neural Circuits* **9**, (2015).
145. Meshalkina, D. A. *et al.* Zebrafish models of autism spectrum disorder. *Experimental Neurology* **299**, 207–216 (2018).
146. Busch-Nentwich, E. The deafness gene *dfna5* is crucial for *ugdh* expression and HA production in the developing ear in zebrafish. *Development* **131**, 943–951 (2004).
147. Gao, M. *et al.* HSF4 regulates lens fiber cell differentiation by activating p53 and its downstream regulators. *Cell Death and Disease* **8**, e3082 (2017).
148. Mishra, S. *et al.* Loss of  $\alpha$ B-crystallin function in zebrafish reveals critical roles in the development of the lens and stress resistance of the heart. *Journal of Biological Chemistry* **293**, 740–753 (2018).
149. Takamiya, M. *et al.* Melanosomes in pigmented epithelia maintain eye lens transparency during zebrafish embryonic development. *Scientific Reports* **6**, (2016).
150. Yousaf, S., Sheikh, S. A., Riazuddin, S., Waryah, A. M. & Ahmed, Z. M. *INPP5K* variant causes autosomal recessive congenital cataract in a Pakistani family. *Clinical Genetics* **93**, 682–686 (2018).
151. Krall, M. *et al.* A zebrafish model of *foxe3* deficiency demonstrates lens and eye defects with dysregulation of key genes involved in cataract formation in humans. *Human Genetics* **137**, 315–328 (2018).
152. Chhetri, J., Jacobson, G. & Gueven, N. Zebrafish—on the move towards ophthalmological research. *Eye* **28**, 367–380 (2014).
153. Evans, P. *et al.* The P23T Cataract Mutation Causes Loss of Solubility of Folded  $\gamma$ D-Crystallin. *Journal of Molecular Biology* **343**, 435–444 (2004).
154. Wu, S.-Y. *et al.* Expression of Cataract-linked  $\gamma$ -Crystallin Variants in Zebrafish Reveals a Proteostasis Network That Senses Protein Stability. *Journal of Biological Chemistry* **291**, 25387–25397 (2016).
155. *Animal models of human disease.* (Elsevier, Acad. Press, 2011).
156. Bibliowicz, J., Tittle, R. K. & Gross, J. M. Toward a Better Understanding of Human Eye Disease. in *Progress in Molecular Biology and Translational Science* vol. 100 287–330 (Elsevier, 2011).
157. Brown, J. D. *et al.* Expression profiling during ocular development identifies 2 *Nlz* genes with a critical role in optic fissure closure. *Proceedings of the National Academy of Sciences* **106**, 1462–1467 (2009).
158. Sanyanusin, P. *et al.* Mutation of the *PAX2* gene in a family with optic nerve colobomas, renal anomalies and vesicoureteral reflux. *Nature Genetics* **9**, 358–364 (1995).
159. Macdonald, R. *et al.* Midline signalling is required for *Pax* gene regulation and patterning of the eyes. *Development* **121**, 3267–3278 (1995).
160. Pais-Roldán, P., Singh, A. P., Schulz, H. & Yu, X. High magnetic field induced otolith fusion in the zebrafish larvae. *Scientific Reports* **6**, (2016).
161. Whitfield, T. T. *et al.* Mutations affecting development of the zebrafish inner ear and lateral line. *Development* **123**, 241–254 (1996).
162. Abbas, L. & Whitfield, T. T. The zebrafish inner ear. in *Fish Physiology* vol. 29 123–171 (Elsevier, 2010).
163. Whitfield, T. T. Zebrafish as a model for hearing and deafness. *Journal of Neurobiology* **53**, 157–171 (2002).
164. Nicolson, T. *et al.* Genetic Analysis of Vertebrate Sensory Hair Cell Mechanosensation: the Zebrafish Circler Mutants. *Neuron* **20**, 271–283 (1998).
165. Armant, O. *et al.* Genome-wide, whole mount in situ analysis of transcriptional regulators in zebrafish embryos. *Developmental Biology* **380**, 351–362 (2013).

166. Asakawa, K. *et al.* Genetic dissection of neural circuits by Tol2 transposon-mediated Gal4 gene and enhancer trapping in zebrafish. *Proceedings of the National Academy of Sciences* **105**, 1255–1260 (2008).
167. Yang, Y. *et al.* PRMT9 is a type II methyltransferase that methylates the splicing factor SAP145. *Nat Commun* **6**, 6428 (2015).
168. Firth, H. V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).
169. Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* **36**, 928–930 (2015).
170. Najmabadi, H. *et al.* Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* **478**, 57–63 (2011).
171. Scheidecker, S. *et al.* Exome sequencing of Bardet–Biedl syndrome patient identifies a null mutation in the BBSome subunit *BBIP1* ( *BBS18* ). *Journal of Medical Genetics* **51**, 132–136 (2014).
172. Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. *Human Mutation* **36**, 928–930 (2015).
173. Stranneheim, H. & Wedell, A. Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders. *Journal of Internal Medicine* **279**, 3–15 (2016).
174. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
175. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297–1303 (2010).
176. Backenroth, D. *et al.* CANOES: detecting rare copy number variants from whole exome sequencing data. *Nucleic Acids Research* **42**, e97–e97 (2014).
177. Geoffroy, V. *et al.* VaRank: a simple and powerful tool for ranking genetic variants. *PeerJ* **3**, e796 (2015).
178. Yeo, G. & Burge, C. B. Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *Journal of Computational Biology* **11**, 377–394 (2004).
179. Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved Splice Site Detection in Genie. *Journal of Computational Biology* **4**, 311–323 (1997).
180. Shapiro, M. B. & Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174 (1987).
181. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
182. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
183. MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L. & Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Research* **42**, D986–D992 (2014).
184. Geoffroy, V. *et al.* AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics* **34**, 3572–3574 (2018).
185. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405–424 (2015).
186. Geoffroy, V. *et al.* Whole-genome sequencing in patients with ciliopathies uncovers a novel recurrent tandem duplication in IFT140. *Human Mutation* **39**, 983–992 (2018).
187. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
188. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

189. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
190. Bartenhagen, C. & Dugas, M. Robust and exact structural variation detection with paired-end and soft-clipped alignments: SoftSV compared with eight algorithms. *Brief. Bioinformatics* **17**, 51–62 (2016).
191. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178–192 (2013).
192. Geoffroy, V. *et al.* VaRank: a simple and powerful tool for ranking genetic variants. *PeerJ* **3**, e796 (2015).
193. Geoffroy, V. *et al.* AnnotSV: An integrated tool for Structural Variations annotation. *Bioinformatics* (2018) doi:10.1093/bioinformatics/bty304.
194. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).
195. Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved splice site detection in Genie. *J. Comput. Biol.* **4**, 311–323 (1997).
196. Shapiro, M. B. & Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174 (1987).
197. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
198. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
199. MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L. & Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* **42**, D986–D992 (2014).
200. O’Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–745 (2016).
201. Tyner, C. *et al.* The UCSC Genome Browser database: 2017 update. *Nucleic Acids Res.* **45**, D626–D634 (2017).
202. Estrada-Cuzcano, A. *et al.* Novel IQCE variations confirm its role in postaxial polydactyly and cause ciliary defect phenotype in zebrafish. *Human Mutation* (2019) doi:10.1002/humu.23924.
203. Mattioli, F. *Dissertation: Identification of novel genetic causes of monogenic intellectual disability.* (2018).
204. Aranda, P. S., LaJoie, D. M. & Jorczyk, C. L. Bleach gel: a simple agarose gel for analyzing RNA quality. *Electrophoresis* **33**, 366–369 (2012).
205. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
206. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
207. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
208. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
209. Wang, L. *et al.* Measure transcript integrity using RNA-seq data. *BMC Bioinformatics* **17**, (2016).
210. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
211. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
212. Hartley, S. W. & Mullikin, J. C. Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. *Nucleic Acids Res* **44**, e127–e127 (2016).
213. Li, Y. I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* **50**, 151–158 (2018).
214. Hartley, S. W. & Mullikin, J. C. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC Bioinformatics* **16**, 224–224 (2015).

215. Stoetzel, C. *et al.* A mutation in VPS15 (PIK3R4) causes a ciliopathy and affects IFT20 release from the cis-Golgi. *Nature Communications* **7**, 13586 (2016).
216. Waltz, F. *et al.* Small is big in Arabidopsis mitochondrial ribosome. *Nat Plants* **5**, 106–117 (2019).
217. Chicois, C. *et al.* The UPF1 interactome reveals interaction networks between RNA degradation and translation repression factors in Arabidopsis. *Plant J* **96**, 119–132 (2018).
218. Westerfield, M. *The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio)*. (University of Oregon Press, Eugene, 2000).
219. Aleström, P. *et al.* Zebrafish: Housing and husbandry recommendations. *Laboratory Animals* 002367721986903 (2019) doi:10.1177/0023677219869037.
220. Gagnon, J. A. *et al.* Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs. *PLoS ONE* **9**, e98186 (2014).
221. Teboul, L., Murray, S. A. & Nolan, P. M. Phenotyping first-generation genome editing mutants: a new standard? *Mammalian Genome* **28**, 377–382 (2017).
222. Küry, S. *et al.* De Novo Disruption of the Proteasome Regulatory Subunit PSMD12 Causes a Syndromic Neurodevelopmental Disorder. *The American Journal of Human Genetics* **100**, 352–363 (2017).
223. Paone, C. *et al.* Loss of zebrafish Smyd1a interferes with myofibrillar integrity without triggering the misfolded myosin response. *Biochemical and Biophysical Research Communications* **496**, 339–345 (2018).
224. Etard, C., Joshi, S., Stegmaier, J., Mikut, R. & Strähle, U. Tracking of Indels by DEcomposition is a Simple and Effective Method to Assess Efficiency of Guide RNAs in Zebrafish. *Zebrafish* **14**, 586–588 (2017).
225. Leventea, E., Hazime, K., Zhao, C. & Malicki, J. Analysis of cilia structure and function in zebrafish. in *Methods in Cell Biology* vol. 133 179–227 (Elsevier, 2016).
226. Berger, J., Sztal, T. & Currie, P. D. Quantification of birefringence readily measures the level of muscle damage in zebrafish. *Biochemical and Biophysical Research Communications* **423**, 785–788 (2012).
227. Rodriguez, A. *et al.* *ToxTrac* : A fast and robust software for tracking organisms. *Methods in Ecology and Evolution* **9**, 460–464 (2018).
228. Rodriguez, A., Zhang, H., Klaminder, J., Brodin, T. & Andersson, M. ToxId: an efficient algorithm to solve occlusions when tracking multiple animals. *Scientific Reports* **7**, (2017).
229. Flint, J. Genetic basis of cognitive disability. *Dialogues Clin Neurosci* **3**, 37–46 (2001).
230. Morales, Y., Cáceres, T., May, K. & Hevel, J. M. Biochemistry and regulation of the protein arginine methyltransferases (PRMTs). *Archives of Biochemistry and Biophysics* **590**, 138–152 (2016).
231. Valente, E. M., Rostj, R. O., Gibbs, E. & Gleeson, J. G. Primary cilia in neurodevelopmental disorders. *Nat Rev Neurol* **10**, 27–36 (2014).
232. Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T. & Saxena, S. Prevalence of intellectual disability: A meta-analysis of population-based studies. *Research in Developmental Disabilities* **32**, 419–436 (2011).
233. Gilissen, C. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344 (2014).
234. Kochinke, K. *et al.* Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. *The American Journal of Human Genetics* **98**, 149–164 (2016).
235. Mehregan, H., Najmabadi, H. & Kahrizi, K. Genetic Studies in Intellectual Disability and Behavioral Impairment. *Arch Iran Med* **19**, 363–375 (2016).
236. Redin, C. *et al.* Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alström Syndromes. *Journal of Medical Genetics* **49**, 502–512 (2012).
237. Tewary, S. K., Zheng, Y. G. & Ho, M.-C. Protein arginine methyltransferases: insights into the enzyme structure and mechanism at the atomic level. *Cellular and Molecular Life Sciences* **76**, 2917–2932 (2019).



238. Bennett, R. L. *et al.* Recommendations for standardized human pedigree nomenclature. Pedigree Standardization Task Force of the National Society of Genetic Counselors. *Am. J. Hum. Genet.* **56**, 745–752 (1995).
239. Harripaul, R. *et al.* Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. *Molecular Psychiatry* **23**, 973–984 (2018).
240. Xu, L.-M. *et al.* AutismKB: an evidence-based knowledgebase of autism genetics. *Nucleic Acids Research* **40**, D1016–D1022 (2012).
241. SYSCILIA Study Group *et al.* The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. *Cilia* **2**, (2013).
242. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).
243. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**, 1–13 (2009).
244. Owlanj, H., Jie Yang, H. & Wei Feng, Z. Nucleoside diphosphate kinase Nm23-M1 involves in oligodendroglial versus neuronal cell fate decision in vitro. *Differentiation* **84**, 281–293 (2012).
245. Funa, K. & Sasahara, M. The Roles of PDGF in Development and During Neurogenesis in the Normal and Diseased Nervous System. *Journal of Neuroimmune Pharmacology* **9**, 168–181 (2014).
246. Dandoy-Dron, F., Griffond, B., Mishal, Z., Tovey, M. G. & Dron, M. Scrg1, a novel protein of the CNS is targeted to the large dense-core vesicles in neuronal cells. *European Journal of Neuroscience* **18**, 2449–2459 (2003).
247. Song, Z., Zhang, X., Jia, S., Yelick, P. C. & Zhao, C. Zebrafish as a Model for Human Ciliopathies. *Journal of Genetics and Genomics* **43**, 107–120 (2016).
248. Vaz, R., Hofmeister, W. & Lindstrand, A. Zebrafish Models of Neurodevelopmental Disorders: Limitations and Benefits of Current Tools and Techniques. *International Journal of Molecular Sciences* **20**, 1296 (2019).
249. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Research* **46**, D754–D761 (2018).
250. Elsen, G. E., Choi, L. Y., Prince, V. E. & Ho, R. K. The autism susceptibility gene met regulates zebrafish cerebellar development and facial motor neuron migration. *Developmental Biology* **335**, 78–92 (2009).
251. Liu, C. *et al.* CRISPR/Cas9-induced shank3b mutant zebrafish display autism-like behaviors. *Molecular Autism* **9**, (2018).
252. Snoeijen-Schouwenaars, F. M. *et al.* Diagnostic exome sequencing in 100 consecutive patients with both epilepsy and intellectual disability. *Epilepsia* **60**, 155–164 (2019).
253. Franić, S. *et al.* Intelligence: shared genetic basis between Mendelian disorders and a polygenic trait. *European Journal of Human Genetics* **23**, 1378–1383 (2015).
254. Akawi, N. *et al.* Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. *Nature Genetics* **47**, 1363–1369 (2015).
255. Wang, Y.-C., Wang, J.-D., Chen, C.-H., Chen, Y.-W. & Li, C. A novel BLAST-Based Relative Distance (BBRD) method can effectively group members of protein arginine methyltransferases and suggest their evolutionary relationship. *Molecular Phylogenetics and Evolution* **84**, 101–111 (2015).
256. Krause, C. D. *et al.* Protein arginine methyltransferases: Evolution and assessment of their pharmacological and therapeutic potential. *Pharmacology & Therapeutics* **113**, 50–87 (2007).
257. Najmabadi, H. *et al.* Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* **478**, 57–63 (2011).
258. Jansen, S. *et al.* De novo variants in FBXO11 cause a syndromic form of intellectual disability with behavioral problems and dysmorphisms. *European Journal of Human Genetics* **27**, 738–746 (2019).
259. Ishikawa, H., Thompson, J., Yates, J. R. & Marshall, W. F. Proteomic Analysis of Mammalian Primary Cilia. *Current Biology* **22**, 414–419 (2012).

260. Narita, K. *et al.* Proteomic analysis of multiple primary cilia reveals a novel mode of ciliary development in mammals. *Biology Open* **1**, 815–825 (2012).
261. Mizuno, K. & Sloboda, R. D. Protein arginine methyltransferases interact with intraflagellar transport particles and change location during flagellar growth and resorption. *Molecular Biology of the Cell* **28**, 1208–1222 (2017).
262. Eguether, T. *et al.* IFT27 Links the BBSome to IFT for Maintenance of the Ciliary Signaling Compartment. *Developmental Cell* **31**, 279–290 (2014).
263. Andreu-Cervera, A. *et al.* The ciliopathy gene *ftm/rpgrrip1* controls mouse forebrain patterning via region-specific modulation of hedgehog/gli signaling. *The Journal of Neuroscience* 2199–18 (2019) doi:10.1523/JNEUROSCI.2199-18.2019.
264. Hirano, S., Yan, Q. & Suzuki, S. T. Expression of a Novel Protocadherin, OL-Protocadherin, in a Subset of Functional Systems of the Developing Mouse Brain. *The Journal of Neuroscience* **19**, 995–1005 (1999).
265. Wang, X. *et al.* Decreased Number and Expression of nNOS-Positive Interneurons in Basolateral Amygdala in Two Mouse Models of Autism. *Frontiers in Cellular Neuroscience* **12**, (2018).
266. Uemura, M., Nakao, S., Suzuki, S. T., Takeichi, M. & Hirano, S. OL-protocadherin is essential for growth of striatal axons and thalamocortical projections. *Nature Neuroscience* **10**, 1151–1159 (2007).
267. Morrow, E. M. *et al.* Identifying Autism Loci and Genes by Tracing Recent Shared Ancestry. *Science* **321**, 218–223 (2008).
268. Zhang, Q., Huang, Y., Zhang, L., Ding, Y.-Q. & Song, N.-N. Loss of *Satb2* in the Cortex and Hippocampus Leads to Abnormal Behaviors in Mice. *Frontiers in Molecular Neuroscience* **12**, (2019).
269. Kozol, R. A. *et al.* Two knockdown models of the autism genes SYNGAP1 and SHANK3 in zebrafish produce similar behavioral phenotypes associated with embryonic disruptions of brain morphogenesis. *Human Molecular Genetics* **24**, 4006–4023 (2015).
270. Di Donato, N. *et al.* Severe forms of Baraitser-Winter syndrome are caused by ACTB mutations rather than ACTG1 mutations. *Eur. J. Hum. Genet.* **22**, 179–183 (2014).
271. Ogris, C., Guala, D. & Sonnhhammer, E. L. L. FunCoup 4: new species, data, and visualization. *Nucleic Acids Res.* **46**, D601–D607 (2018).
272. Myhre, J. L., Hills, J. A., Jean, F. & Pilgrim, D. B. *Unc45b* is essential for early myofibrillogenesis and costamere formation in zebrafish. *Developmental Biology* **390**, 26–40 (2014).
273. Plaisancié, J. *et al.* Genetics of anophthalmia and microphthalmia. Part 1: Non-syndromic anophthalmia/microphthalmia. *Human Genetics* **138**, 799–830 (2019).
274. FitzPatrick, D. R. & Heyningen, V. van. Developmental eye disorders. *Current Opinion in Genetics & Development* **15**, 348–353 (2005).
275. Ma, X., Kawamoto, S., Hara, Y. & Adelstein, R. S. A Point Mutation in the Motor Domain of Nonmuscle Myosin II-B Impairs Migration of Distinct Groups of Neurons. *Molecular Biology of the Cell* **15**, 2568–2579 (2004).
276. Ma, X., Kawamoto, S., Uribe, J. & Adelstein, R. S. Function of the Neuron-specific Alternatively Spliced Isoforms of Nonmuscle Myosin II-B during Mouse Brain Development. *Molecular Biology of the Cell* **17**, 2138–2149 (2006).
277. Verloes, A. *et al.* Baraitser–Winter cerebrofrontofacial syndrome: delineation of the spectrum in 42 cases. *European Journal of Human Genetics* **23**, 292–301 (2015).
278. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
279. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248–249 (2010).
280. Ricketson, D., Johnston, C. A. & Prehoda, K. E. Multiple tail domain interactions stabilize nonmuscle myosin II bipolar filaments. *Proceedings of the National Academy of Sciences* **107**, 20964–20969 (2010).
281. Wilson, C. A. *et al.* Myosin II contributes to cell-scale actin network treadmill through network disassembly. *Nature* **465**, 373–377 (2010).

282. Gutzman, J. H., Sahu, S. U. & Kwas, C. Non-muscle myosin IIA and IIB differentially regulate cell shape changes during zebrafish brain morphogenesis. *Developmental Biology* **397**, 103–115 (2015).
283. Lahrouchi, N. *et al.* Homozygous frameshift mutations in FAT1 cause a syndrome characterized by colobomatous-microphthalmia, ptosis, nephropathy and syndactyly. *Nature Communications* **10**, (2019).
284. Yates, T. M., Turner, C. L., Firth, H. V., Berg, J. & Pilz, D. T. Baraitser-Winter cerebrofrontofacial syndrome: Baraitser-Winter cerebrofrontofacial syndrome. *Clinical Genetics* **92**, 3–9 (2017).
285. Rivière, J.-B. *et al.* De novo mutations in the actin genes ACTB and ACTG1 cause Baraitser-Winter syndrome. *Nature Genetics* **44**, 440–444 (2012).
286. Okumura, T. *et al.* Association of PAX2 and Other Gene Mutations with the Clinical Manifestations of Renal Coloboma Syndrome. *PLOS ONE* **10**, e0142843 (2015).
287. Conti, M. A., Even-Ram, S., Liu, C., Yamada, K. M. & Adelstein, R. S. Defects in Cell Adhesion and the Visceral Endoderm following Ablation of Nonmuscle Myosin Heavy Chain II-A in Mice. *Journal of Biological Chemistry* **279**, 41263–41266 (2004).
288. Bao, J., Jana, S. S. & Adelstein, R. S. Vertebrate Nonmuscle Myosin II Isoforms Rescue Small Interfering RNA-induced Defects in COS-7 Cell Cytokinesis. *Journal of Biological Chemistry* **280**, 19594–19599 (2005).
289. Ma, X. & Adelstein, R. S. The role of vertebrate nonmuscle Myosin II in development and human disease. *BioArchitecture* **4**, 88–102 (2014).
290. Vicente-Manzanares, M., Koach, M. A., Whitmore, L., Lamers, M. L. & Horwitz, A. F. Segregation and activation of myosin IIB creates a rear in migrating cells. *The Journal of Cell Biology* **183**, 543–554 (2008).
291. Tuzovic, L. *et al.* A human de novo mutation in MYH10 phenocopies the loss of function mutation in mice. *Rare Diseases* **1**, e26144 (2013).
292. Hamdan, F. F. *et al.* De Novo Mutations in Moderate or Severe Intellectual Disability. *PLoS Genetics* **10**, e1004772 (2014).
293. Kim, H.-T. *et al.* Myh10 deficiency leads to defective extracellular matrix remodeling and pulmonary disease. *Nature Communications* **9**, (2018).
294. Pollard, T. D. & Borisy, G. G. Cellular Motility Driven by Assembly and Disassembly of Actin Filaments. *Cell* **112**, 453–465 (2003).
295. Veigel, C. & Schmidt, C. F. Moving into the cell: single-molecule studies of molecular motors in complex environments. *Nature Reviews Molecular Cell Biology* **12**, 163–176 (2011).
296. Haviv, L., Gillo, D., Backouche, F. & Bernheim-Groswasser, A. A Cytoskeletal Demolition Worker: Myosin II Acts as an Actin Depolymerization Agent. *Journal of Molecular Biology* **375**, 325–330 (2008).
297. MacArthur, M. W. & Thornton, J. M. Influence of proline residues on protein conformation. *Journal of Molecular Biology* **218**, 397–412 (1991).
298. Schimmel, P. R. & Flory, P. J. Conformational energies and configurational statistics of copolypeptides containing l-proline. *Journal of Molecular Biology* **34**, 105–120 (1968).
299. Yang, Q., Zhang, X.-F., Pollard, T. D. & Forscher, P. Arp2/3 complex-dependent actin networks constrain myosin II function in driving retrograde actin flow. *The Journal of Cell Biology* **197**, 939–956 (2012).
300. Niceta, M. *et al.* Mutations Impairing GSK3-Mediated MAF Phosphorylation Cause Cataract, Deafness, Intellectual Disability, Seizures, and a Down Syndrome-like Facies. *The American Journal of Human Genetics* **96**, 816–825 (2015).
301. De Franco, E. *et al.* Dominant ER Stress-Inducing *WFS1* Mutations Underlie a Genetic Syndrome of Neonatal/Infancy-Onset Diabetes, Congenital Sensorineural Deafness, and Congenital Cataracts. *Diabetes* **66**, 2044–2053 (2017).
302. Reis, L. M. & Semina, E. V. Genetic landscape of isolated pediatric cataracts: extreme heterogeneity and variable inheritance patterns within genes. *Human Genetics* (2018) doi:10.1007/s00439-018-1932-x.

303. Azaiez, H. *et al.* Genomic Landscape and Mutational Signatures of Deafness-Associated Genes. *The American Journal of Human Genetics* **103**, 484–497 (2018).
304. TANAKA, K. The proteasome: Overview of structure and functions. *Proceedings of the Japan Academy, Series B* **85**, 12–36 (2009).
305. Steffen, J., Seeger, M., Koch, A. & Kruger, E. Proteasomal degradation is transcriptionally controlled by TCF11 via an ERAD-dependent feedback loop. *Molecular cell* **40**, 147–58 (2010).
306. Radhakrishnan, S. K. *et al.* Transcription factor Nrf1 mediates the proteasome recovery pathway after proteasome inhibition in mammalian cells. *Molecular cell* **38**, 17–28 (2010).
307. Sotzny, F. *et al.* TCF11/Nrf1-Mediated Induction of Proteasome Expression Prevents Cytotoxicity by Rotenone. *Antioxidants & redox signaling* **25**, 870–885 (2016).
308. Thisse, B. & Thisse, C. Fast Release Clones: A High Throughput Expression Analysis. *ZFIN Direct Data Submission: (Unpublished)* (2004).
309. Geng, F.-S. *et al.* Semicircular canal morphogenesis in the zebrafish inner ear requires the function of gpr126 (lauscher), an adhesion class G protein-coupled receptor gene. *Development* **140**, 4362–4374 (2013).
310. Han, Y. *et al.* Grhl2 deficiency impairs otic development and hearing ability in a zebrafish model of the progressive dominant hearing loss DFNA28. *Human Molecular Genetics* **20**, 3213–3226 (2011).
311. Stooke-Vaughan, G. A., Obholzer, N. D., Baxendale, S., Megason, S. G. & Whitfield, T. T. Otolith tethering in the zebrafish otic vesicle requires Otogelin and -Tectorin. *Development* **142**, 1137–1145 (2015).
312. Elsen, G. E., Choi, L. Y., Prince, V. E. & Ho, R. K. The autism susceptibility gene met regulates zebrafish cerebellar development and facial motor neuron migration. *Developmental Biology* **335**, 78–92 (2009).
313. Kindt, K. S., Finch, G. & Nicolson, T. Kinocilia Mediate Mechanosensitivity in Developing Zebrafish Hair Cells. *Developmental Cell* **23**, 329–341 (2012).
314. Belkadi, A. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A* **112**, 5473–5478 (2015).
315. Vaz-Drago, R., Custódio, N. & Carmo-Fonseca, M. Deep intronic mutations and human disease. *Human Genetics* **136**, 1093–1111 (2017).
316. Husnjak, K. *et al.* Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* **453**, 481–488 (2008).
317. Deveraux, Q., Ustrell, V., Pickart, C. & Rechsteiner, M. A 26 S protease subunit that binds ubiquitin conjugates. *Journal of Biological Chemistry* **269**, 7059–7061 (1994).
318. Yao, T. & Cohen, R. E. A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* **419**, 403–407 (2002).
319. Chen, S. *et al.* Structural basis for dynamic regulation of the human 26S proteasome. *Proceedings of the National Academy of Sciences* **113**, 12991–12996 (2016).
320. Lam, Y. A., Lawson, T. G., Velayutham, M., Zweier, J. L. & Pickart, C. M. A proteasomal ATPase subunit recognizes the polyubiquitin degradation signal. *Nature* **416**, 763–767 (2002).
321. Khalil, R. *et al.* PSMD12 haploinsufficiency in a neurodevelopmental disorder with autistic features. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **177**, 736–745 (2018).
322. Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 531210 (2019) doi:10.1101/531210.
323. Huang, N., Lee, I., Marcotte, E. M. & Hurles, M. E. Characterising and Predicting Haploinsufficiency in the Human Genome. *PLOS Genetics* **6**, e1001154 (2010).
324. Ma, C. P., Slaughter, C. A. & DeMartino, G. N. Identification, purification, and characterization of a protein activator (PA28) of the 20 S proteasome (macropain). *Journal of Biological Chemistry* **267**, 10515–10523 (1992).

325. Yang, K., Huang, R., Fujihira, H., Suzuki, T. & Yan, N. N-glycanase NGLY1 regulates mitochondrial homeostasis and inflammation through NRF1. *The Journal of experimental medicine* **215**, 2600–2616 (2018).
326. Widenmaier, S. B. *et al.* NRF1 Is an ER Membrane Sensor that Is Central to Cholesterol Homeostasis. *Cell* **171**, 1094–1109 e15 (2017).
327. Agarwal, A. K. *et al.* PSMB8 encoding the beta5i proteasome subunit is mutated in joint contractures, muscle atrophy, microcytic anemia, and panniculitis-induced lipodystrophy syndrome. *American journal of human genetics* **87**, 866–72 (2010).
328. Liu, Y. *et al.* Mutations in proteasome subunit beta type 8 cause chronic atypical neutrophilic dermatosis with lipodystrophy and elevated temperature with evidence of genetic and phenotypic heterogeneity. *Arthritis and rheumatism* **64**, 895–907 (2012).
329. Arima, K. *et al.* Proteasome assembly defect due to a proteasome subunit beta type 8 (PSMB8) mutation causes the autoinflammatory disorder, Nakajo-Nishimura syndrome. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 14914–9 (2011).
330. Brehm, A. *et al.* Additive loss-of-function proteasome subunit mutations in CANDLE/PRAAS patients promote type I IFN production. *The Journal of clinical investigation* **125**, 4196–211 (2015).
331. Poli, M. C. *et al.* Heterozygous Truncating Variants in POMP Escape Nonsense-Mediated Decay and Cause a Unique Immune Dysregulatory Syndrome. *American journal of human genetics* **102**, 1126–1142 (2018).
332. Shang, F. & Taylor, A. Chapter 10 - Role of the Ubiquitin-Proteasome in Protein Quality Control and Signaling: Implication in the Pathogenesis of Eye Diseases. in *Progress in Molecular Biology and Translational Science* (ed. Grune, T.) vol. 109 347–396 (Academic Press, 2012).
333. Kishino, T., Lalonde, M. & Wagstaff, J. UBE3A/E6-AP mutations cause Angelman syndrome. *Nature Genetics* **15**, 70–73 (1997).
334. Nascimento, R. M. P., Otto, P. A., de Brouwer, A. P. M. & Vianna-Morgante, A. M. UBE2A, Which Encodes a Ubiquitin-Conjugating Enzyme, Is Mutated in a Novel X-Linked Mental Retardation Syndrome. *The American Journal of Human Genetics* **79**, 549–555 (2006).
335. Froyen, G. *et al.* Submicroscopic Duplications of the Hydroxysteroid Dehydrogenase HSD17B10 and the E3 Ubiquitin Ligase HUWE1 Are Associated with Mental Retardation. *The American Journal of Human Genetics* **82**, 432–443 (2008).
336. Moortgat, S. *et al.* HUWE1 variants cause dominant X-linked intellectual disability: a clinical study of 21 patients. *Eur J Hum Genet* **26**, 64–74 (2018).
337. Santiago-Sim, T. *et al.* Biallelic Variants in OTUD6B Cause an Intellectual Disability Syndrome Associated with Seizures and Dysmorphic Features. *The American Journal of Human Genetics* **100**, 676–688 (2017).
338. Imai, F., Yoshizawa, A., Fujimori-Tonou, N., Kawakami, K. & Masai, I. The ubiquitin proteasome system is required for cell proliferation of the lens epithelium and for differentiation of lens fiber cells in zebrafish. *Development* **137**, 3257–3268 (2010).
339. Millimaki, B. B., Sweet, E. M., Dhasan, M. S. & Riley, B. B. Zebrafish *atoh1* genes: classic proneural activity in the inner ear and regulation by Fgf and Notch. *Development* **134**, 295–305 (2007).
340. Cruz, S., Shiao, J.-C., Liao, B.-K., Huang, C.-J. & Hwang, P.-P. Plasma membrane calcium ATPase required for semicircular canal formation and otolith growth in the zebrafish inner ear. *Journal of Experimental Biology* **212**, 639–647 (2009).
341. Haddon, C. M. & Lewis, J. H. Hyaluronan as a propellant for epithelial movement: the development of semicircular canals in the inner ear of *Xenopus*. *Development* **112**, 541–550 (1991).
342. Scott, A. F., Bocchini, C. A., Amberger, J. S. & Hamosh, A. OMIM.org: leveraging knowledge across phenotype–gene relationships. *Nucleic Acids Research* **47**, D1038–D1043 (2018).
343. Ogris, C., Guala, D., Kaduk, M. & Sonhammer, E. L. L. FunCoup 4: new species, data, and visualization. *Nucleic Acids Research* **46**, D601–D607 (2017).
344. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).

345. Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved Splice Site Detection in Genie. *Journal of Computational Biology* **4**, 311–323 (1997).
346. Shapiro, M. B. & Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174 (1987).
347. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073–1081 (2009).
348. Geoffroy, V. *et al.* VaRank: a simple and powerful tool for ranking genetic variants. *PeerJ* **3**, e796 (2015).
349. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **45**, D158–D169 (2016).
350. Binato, R., Alvarez Martinez, C. E., Pizzatti, L., Robert, B. & Abdelhay, E. SMAD 8 binding to mice *Msx1* basal promoter is required for transcriptional activation. *Biochem J* **393**, 141 (2006).
351. Guerrero, C., Milenković, T., Pržulj, N., Kaiser, P. & Huang, L. Characterization of the proteasome interaction network using a QTAX-based tag-team strategy and protein interaction network analysis. *Proc Natl Acad Sci USA* **105**, 13333 (2008).
352. Wang, Y. *et al.* Coiled-coil networking shapes cell molecular machinery. *MBoC* **23**, 3911–3922 (2012).
353. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
354. Schwarze, K., Buchanan, J., Taylor, J. C. & Wordsworth, S. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genetics in Medicine* **20**, 1122–1130 (2018).
355. Thuresson, A. *et al.* Whole genome sequencing of consanguineous families reveals novel pathogenic variants in intellectual disability. *Clinical Genetics* **95**, 436–439 (2019).
356. Berberich, A. J., Ho, R. & Hegele, R. A. Whole genome sequencing in the clinic: empowerment or too much information? *Canadian Medical Association Journal* **190**, E124–E125 (2018).
357. Chrystoja, C. C. & Diamandis, E. P. Whole Genome Sequencing as a Diagnostic Test: Challenges and Opportunities. *Clinical Chemistry* **60**, 724–733 (2014).
358. Abramowicz, A. & Gos, M. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of Applied Genetics* **59**, 253–268 (2018).
359. Vaz-Drago, R., Custódio, N. & Carmo-Fonseca, M. Deep intronic mutations and human disease. *Human Genetics* **136**, 1093–1111 (2017).
360. El-Brolosy, M. A. & Stainier, D. Y. R. Genetic compensation: A phenomenon in search of mechanisms. *PLOS Genetics* **13**, e1006780 (2017).
361. Stainier, D. Y. R. *et al.* Guidelines for morpholino use in zebrafish. *PLOS Genetics* **13**, e1007000 (2017).
362. Hug, N., Longman, D. & Cáceres, J. F. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research* **44**, 1483–1495 (2016).
363. El-Brolosy, M. A. *et al.* Genetic compensation triggered by mutant mRNA degradation. *Nature* **568**, 193–197 (2019).
364. Fleisch, V. C. & Neuhauss, S. C. F. Visual behavior in zebrafish. *Zebrafish* **3**, 191–201 (2006).
365. Kremer, L. S. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nature Communications* **8**, (2017).

# Curriculum Vitae (CV)

## Personal Data

Name: Ariane Kröll-Hermi  
Date and place of birth: 03.06.1991  
Nationality: German

## PhD

Since 10/2016 **PhD,**  
Laboratory of Medical Genetics, Université de Strasbourg, France  
Institute of Toxicology and Genetics, Karlsruher Institut für  
Technologie, Germany

## Studies

09/2015-11/2016 **Master of Science,** Studies in Biology and Molecular Genetics  
Université de Strasbourg, Strasbourg, France  
10/2014-06/2016 **Master of Science,** Studies in Human- and Molecular Biology  
Universität des Saarlandes, Saarbrücken, Germany  
04/2011-05/2014 **Bachelor of Science,** Studies in Biology  
Johannes Gutenberg-Universität, Mainz, Germany

## School Education

04/2011 – 05/2014 **A-levels**  
Gymnasium an der Stadtmauer, Bad Kreuznach, Germany

# Publications

## Published:

- Véronique Geoffroy, Corinne Stoetzel, Sophie Scheidecker, Elise Schaefer, Isabelle Perrault, Séverine Bär, **Ariane Kröll**, et al. (2018) Whole-genome sequencing in patients with ciliopathies uncovers a novel recurrent tandem duplication in IFT140. *Human Mutation*. Volume 39, Issue 7

## Submitted publications:

- **Ariane Kröll-Hermi\***, Sophie Scheidecker\*, Séverine Bär\*, Jean Muller, Véronique Geoffroy, Elise Schaefer, Stéphane Kremer, Christelle Etard, Corinne Stoetzel, Uwe Strähle, Sylvie Friant<sup>#</sup> and Hélène Dollfus<sup>#</sup> (2019) “*MYH10* dominant variant causes Baraitser-Winter cerebro-fronto-facial syndrome related phenotype with major ophthalmic developmental features.”  
(\* joint first authors, # equal contributors)  
**[Manuscript submitted to the “Journal of Medical Genetics”]**
- **Ariane Kröll-Hermi\***, Frédéric Ebstein\*, Corinne Stoetzel\*, Véronique Geoffroy\*, Elise Schaefer, Sophie Scheidecker, Séverine Bär, Masanari Takamiya, Koichi Kawakami, Barbara A. Zieba, Fouzia Studer, Valerie Pelletier, Claude Speeg-Schatz, Vincent Laugel, Dan Lipsker, Florian Sandron, Steven McGinn, Anne Boland, Jean-François Deleuze, Lauriane Kuhn, Johana Chicher, Philippe Hammann, Sylvie Friant, Christelle Etard, Elke Krüger\*, Jean Muller\*, Uwe Strähle\*, Hélène Dollfus\* (2019) “Proteasome subunit PSMC3 variants cause neurosensory syndrome combining deafness and cataract due to proteotoxic stress.”  
(\* equal contributors)  
**[Manuscript submitted to the journal “EMBO Molecular Medicine”]**

## Publications in preparation:

- **Ariane Kröll-Hermi**, Corinne Stoetzel, Elise Schaefer, Sophie Scheidecker, Megana Prasad, Christelle Etard, Kimia Kahrizi, Véronique Geoffroy, Aline Schneider, Francesca Mattioli, Damien Plassard, Céline Keime, Amélie Piton, Ingrid Wentzensen, Teresa Santiago-Sim, Cecilie Rustad, B. Oyvind, Kristian Tveten, Marie Smeland Falkenberg, Ebba Alkhunaizi, Chumei Li, Susanne Kircher, Franco Laconne, Caroline Maria Kolvenbach, Velibor Tasic, Friedhelm Hildebrandt, Boris Keren, Sandra Whalen, Joelle Roume, Arie van Haeringen, Hossein Najmabadi, Uwe Strähle, Hélène Dollfus and Jean Muller (2019) “Biallelic variations in *PRMT9* delineate a novel syndromic form of intellectual.”  
**[Manuscript in preparation, will be submitted to the “American Journal of Human Genetics”.]**



## Talks/ Poster presentations

### Talk:

- Talk: Deep intronic mutations unveils a novel syndrome.  
The 2nd GENMED Rare Diseases Workshop, 20th of May 2019, CNRGH, Évry

### Poster:

- Poster: Zebrafish, as a useful model to validate human eye candidate diseases genes.  
2019 ISGEDR Meeting – 29th to 31<sup>st</sup> August, Justus-Liebig-University, Giessen, Germany

## Identification and Validation of novel genes implicated in Neurosensory and Neurological Diseases

### Résumé

Dans le monde, on dénombre plus de 7000 maladies rares pour lesquelles 80 % sont d'origines génétiques et pour lesquelles le ou les gènes ne sont pas encore tous connus. Bien que l'intérêt pour les maladies rares ait augmenté au cours des dernières années, il n'existe souvent aucun traitement pour celles-ci. Au cours de ma thèse, je me suis concentré sur l'identification et la validation de nouveaux gènes associés à des maladies neurosensorielles et neurodéveloppementales. La stratégie que j'ai utilisée au cours de ma thèse afin d'identifier de nouveaux gènes responsables de maladies repose principalement sur l'analyse des données de séquençage à haut débit de patients, la confirmation des variants par séquençage Sanger et la validation des gènes candidats en utilisant les cellules des patients et le modèle du poisson zèbre. Dans l'ensemble, j'ai contribué à identifier trois nouveaux gènes associés aux maladies neurodéveloppementales et neurosensorielles. Les résultats obtenus dans cette thèse améliorent la compréhension de la physiopathologie de ces maladies et pourraient aider à trouver de nouvelles cibles thérapeutiques.

**Mot clés :** NGS, maladies neurodéveloppementales, maladies neurosensorielles, *PRMT9*, *MYH10*, *PSMC3*

### Summary

Worldwide, there are more than 7,000 rare diseases for which 80 % are of genetic origin and for which the gene or genes are not yet all known. Although the interest in rare diseases has increased in recent years, there is a lack of therapeutic strategies for most of them. During my thesis, I focused on the identification and validation of new genes associated with neurosensory and neurodevelopmental diseases. The strategy I used during my thesis to identify new disease-causing genes is based primarily on the analysis of high-throughput sequencing data from patients, the confirmation of variants by Sanger sequencing, and the performance of functional experiments using patient cells and the zebrafish model to proof the pathogenicity of the novel gene mutations. Overall, I contributed to the identification of three novel disease genes associated with neurodevelopmental and neurosensory diseases. The results obtained in this thesis improve the understanding of the pathophysiology of these diseases and may help to find new therapeutic targets.

**Key words:** NGS, neurodevelopmental diseases, neurosensory diseases, *PRMT9*, *MYH10*, *PSMC3*